

1-20-2020

Exploring the Neurocognitive Bases of Statistical Learning

Oliver M. Sawi

University of Connecticut - Storrs, oliver.sawi@uconn.edu

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Sawi, Oliver M., "Exploring the Neurocognitive Bases of Statistical Learning" (2020). *Doctoral Dissertations*. 2410.

<https://opencommons.uconn.edu/dissertations/2410>

Exploring the Neurocognitive Bases of Statistical Learning

Oliver M. Sawi, PhD

University of Connecticut, 2020

Statistical Learning (SL) involves the extraction of organizing principles from a set of inputs. Recent advances in SL suggest that SL is a componential construct. To better characterize the componential nature of SL, a strategy may be to turn to literature regarding memory and learning. The current study sought to extend the literature by further characterizing the componential nature of. Aim 1 examined the effect of instruction type (explicit, implicit) on direct and indirect (explicit, implicit) indices of visual statistical learning (VSL) performance. Several studies have suggested explicit instructions shift engagement of additional explicit memory resources improving performance. There were no differences in indirect or direct measures of VSL performance. However, the relationship between direct and indirect measures of VSL was affected by instructional condition suggesting the processes underlying VSL may have been affected. Aim 2 examined the relationship between VSL performance and implicit and explicit memory/learning. Further, Aim 2 examined whether instructional condition affected the relationship between VSL and multiple memory systems. The relationship between the direct measure of VSL and explicit and implicit memory was inconsistent. However, the direct measure patterned similarly across explicit and implicit memory (positive relationship, not affected by instructional condition). The relationship between the indirect measures of VSL and multiple memory systems was similarly inconsistent, but had a similar patterning in the significant cases (positive relationship in the explicit condition,

but negative relationship in the implicit condition). This suggests the indirect measure of VSL was affected by the instructional condition to differentially emphasize aspects of memory systems. In addition, In recent years, several methodological issues have been identified regarding measures established in the literature. To address the inconsistencies in the findings and these concerns, the psychometric properties of the established measures were examined. Exploratory Aim 3 sought to improve upon the processing of these measures using advanced statistical methods and provide recommendations regarding best practices for individual differences analyses. In Exploratory Aim 4, the first set of results were revisited in an exploratory manner using the insights gained from the updated measures. Implications for the characterization of the componential nature of SL were discussed.

Exploring the Neurocognitive Bases of Statistical Learning

Oliver M. Sawi

M.S., University of Connecticut, 2016

B.A., San Francisco State University, 2013

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2020

Copyright by

Oliver Sawi

2020

ii

APPROVAL PAGE

Doctor of Philosophy Dissertation

Exploring the Neurocognitive Bases of Statistical Learning

Presented by

Oliver Sawi, M.S.

Major Advisor _____
Jay Rueckl

Associate Advisor _____
Kenneth Pugh

Associate Advisor _____
Devin Kearns

University of Connecticut

2020

Acknowledgements

I would like to express my gratitude to my primary advisor, Jay Rueckl, for his support and critical feedback on all aspects of this thesis. Without his guidance this project would not have been possible. I would also like to thank my committee members, Ken Pugh and Devin Kearns, for their constant support, guidance, and encouragement; my wonderful mentors such as Fumiko Hoeft who challenged me to connect theory to practice in a meaningful way; and to my research assistants for their dedication to collecting high quality data.

I am deeply grateful to my wife, Huong Nguyen, and her family for their support and understanding throughout this process. Thank you for always supporting me in all aspects of life and never letting me give up. To my own family, thank you for your patience with me and the love you sent me as I moved more than 3,000 miles away, while not being able to return home for years at a time.

Lastly, to my community. Thank you for lifting me up. Thank you for helping me find balance and peace. Thank you for allowing me to represent our community. To be the first was an honor and a burden all at once, but with your support I was able to thrive. Special thanks to Angela Rola and Sheila Kucko from the Asian American Cultural Center and Fany Hannon and Jen Morenus from the Puerto Rican/Latin American Cultural Center.

Introduction

Statistical learning (SL) involves the extraction of the organizing principles or regularities from a set of inputs (e.g., Frost et al., 2015; Siegelman et al., 2017). While there is general agreement that individuals show sensitivity to the statistical properties of inputs, the exact nature of the underlying processes supporting SL are not well understood. The lack of understanding regarding the processes supporting SL has produced inconsistent theories regarding its connection to potentially related constructs such as memory and executive function. Early descriptions of SL typically assumed SL was a unitary, domain general learning mechanism or capacity (Kirkham et al., 2002; Saffran, 2003). Siegelman et al. (2017) noted that most examinations do not describe specific underlying computations or mechanisms, but rather more abstracted systems in which a unified capacity is controlled by a single learning system across all domains. However, recent evidence suggests that SL is in fact componential, suggesting SL is instead a construct comprised of several aspects that may be supported by disparate cognitive systems (see Siegelman et al., 2017; 2015; Arciuli, 2017 for discussion). For example, individual differences do not correlate across modalities using variants of the same canonical SL task (Siegelman et al., 2015). If SL was a unitary learning mechanism that was the same across modalities, one would expect performance on visual and auditory forms of SL to correlate. In addition, in some forms of SL, there is no cross-modality interference but strong inter-modality interference (Conway and Christiansen, 2006) and learning does not transfer across modalities (Redington and Chater, 1996). If SL was a unitary learning mechanism that was activated regardless of modality, there

should be both cross and inter-modality interferences and learning should transfer across modalities.

Due to these findings, Frost et al. (2015) posited a framework for describing the underlying mechanisms of SL as a set of interrelated and modality (e.g., visual, auditory, tactile) specific processes. Their framework posits computations and neurobiological networks are initially constrained by modality followed by a system of domain general computations (or “principles”). Overall, SL paradigms activate higher order processing in networks associated with each specific modality. For example, in word segmentation of continuous speech, aspects of the auditory network are activated such as the inferior frontal gyrus and left temporal gyrus (Karuza et al., 2013; Alba and Okanoya, 2008). Relatedly, visual SL with shapes activates aspects of visual networks (Turk-Browne et al., 2009; Bishoff-Grethe et al., 2000). Further, tasks involving motor involvement such as the serial reaction time task (SRT) produce activation in parietal cortices, motor cortices, and the cerebellum (Packard and Knowlton, 2002). Therefore, at least some of the issues with understanding SL as a unitary learning system can be explained by early modality specific activity (Frost et al., 2015).

While there is strong evidence for modality specific neural processing, there is also neurocognitive evidence suggesting domain general patterns of activation which govern learning of statistical regularities. Within their framework, Frost et al. (2015) posit these domain general principles may emerge in two ways (Frost et al., 2015). First, across modality (e.g., visual, auditory) similar computations are engaged to pull out statistical regularities in the input stream (as modeled by Thiessen et al., 2013; 2015). Second, modality specific information (e.g., representations) generated during initial encoding is

further processed in multi-modal regions. Information across all domains is therefore processed in the same brain networks and may be subject to similar processing demands. Specifically, these multi-modal processing regions include aspects of the frontal (Karuza et al., 2013; Alba and Okanoya, 2008), striatal (Turk-Browne et al., 2009), and Medial Temporal Lobe (MTL) memory systems (Turk-Browne et al., 2009; Schapiro et al., 2015).

Relatedly, theories of componentiality in SL are not only driven by constraints related to differences in modality. Rather, componentiality may emerge from other factors such as differences in aspects of memory or the computations underlying SL. Arciuli (2017) suggested SL is supported by additional components related to encoding and retention. For example, older children performed better than younger children on an SL task, regardless of differences in attention (Arciuli and Simpson, 2012). They posited that an implicit form of working memory (WM) is an underlying component of SL that is late developing, contributing to these age-related differences. In addition, Arciuli (2017) describe an opposite set of findings in Jeste et al. (2015). Arciuli (2017) suggested that the opposite pattern found in a separate study by Jeste et al. (2015) is due to differences in encoding, retention, and development of the systems underlying these processes. Arciuli and Simpson (2012) and Jeste et al. (2015) differ in terms of the age of participants (ages 5-12 and 2-6 respectively), complexity of the sequences used, and familiarity of the encoded object. Arciuli (2017) posited the more complex sequences in Arciuli and Simpson (2012) are more taxing for an aspect of implicit WM which develops slowly. As the sequences in Jeste et al. (2015) were not complex enough, implicit WM was not engaged and therefore there were no differences due to age. In addition, Jeste et al. (2015) used familiar objects, while Arciuli and Simpson (2012) used unrecognizable

shapes, suggesting differential engagement of memory systems during encoding and retention. Taken together, some components of SL are affected by memory-based functions and have differential developmental time scales (e.g., later development of WM) supported by different task-dependent constraints (Arciuli, 2017).

In conclusion, recent advances in SL suggest that SL is a componential construct with components related to modality, encoding, and retention (Frost et al., 2015; Arciuli, 2017). SL is driven and constrained by modality specific factors and domain general principles (Frost et al., 2015). However, although several recent theoretical frameworks have explored the componential nature of SL, important questions remain particularly with regard to the multimodal processing systems (MTL, striatum, neocortex) supporting SL. To better understand the domain-general principles arising from multimodal processing regions (e.g., MTL, striatum, frontal) and processes supporting encoding and retention of SL, a strategy may be to turn to literature regarding memory and learning. For example, the neural correlates of SL overlap with the neural correlates of multiple memory systems. In addition, aspects of encoding and retention affect SL. Grounding SL in more well-established memory theories, which have a long history of exploring the componential nature of memory systems, can help provide insight into the componential nature of SL. As such, a mechanistic description of SL processes should be grounded in more well-established theoretical frameworks and research traditions regarding learning and memory systems.

Multiple Memory Systems Frameworks

Memory is comprised of several distinguishable component processes which support different types of information (Schacter, 1987). There are several frameworks

used to dichotomize memory of these types. For example, declarative and procedural memory (e.g., Squire, 1992; 2004; Ullman, 2004) are typically characterized by dependence on specific anatomical regions such as the MTL or the striatum and neocortex respectively (e.g., Squire, 1992; 2004). In addition, declarative memory is typically associated with memories to which individuals have conscious access such as memories of facts (“semantic”) or events (“episodic”). Procedural memory refers to memories to which individuals do not have conscious access such as skills and habits (e.g., Squire, 1992; 2004). Learning in declarative memory typically occurs with conscious intention while learning in procedural memory occurs over time without direct conscious awareness or intention. Declarative memory is also responsible for the learning arbitrary relationships (associative binding) over short periods of time and is domain general (Cohen et al., 1997; Eichenbaum and Cohen, 2001; Squire and Knowlton, 2000). Learning in procedural memory is related to the understanding of relationships between complex sequences (sensorimotor or cognitive) over extended periods of time and is modality specific (Ullman, 2004; Squire and Knowlton, 2000).

A similar and overlapping dichotomy is the distinction between implicit and explicit memory (e.g., Schacter, 1987). This dichotomy differs from the declarative/procedural distinction in relative focus on the intention of retrieval of information, rather than neural correlates. For example, explicit memory refers to intentional and conscious retrieval of information. It is often measured by direct means such as recall or recognition, and it depends on attentional control (executive function) and working memory (e.g., DeKeyser, 2003; Liu et al., 2015). The implicit memory system on the other hand, refers to incidental and unconscious retrieval of information. Explicit and implicit memory systems map onto

similar neural correlates as do declarative and procedural memory systems . The explicit memory system is typically characterized by the reliance on the MTL system, while the implicit memory system seems to rely on a circuit including frontal-striatal connections (Dew and Cabeza, 2011; Voss and Paller, 2008).

In addition, there is a dichotomy related to explicit and implicit memory called explicit and implicit *learning* (Reber, 1992; 2003). *Learning* typically refers to the process by which information is generated (e.g., incidental or intentional, unconscious or conscious) rather than the process by which information is retrieved. Examinations of implicit and explicit learning are typically concerned with whether learning occurred in an incidental, unconscious fashion (e.g., Yang and Li, 2012; Destrebecqz et al., 2001, 2005). Conversely, examinations of explicit and implicit memory are concerned with whether the knowledge accrued reflects conscious retrieval. They are similarly concerned with the type of measurement, direct (e.g., recognition) or indirect (e.g., reaction time), which is related to explicit and implicit memory respectively (e.g., Voss and Paller, 2008). Implicit learning refers to learning in incidental, unconscious conditions over time, while explicit learning refers to learning in deliberate, conscious conditions and can occur over short periods of time. In addition, some studies suggest working memory is related to explicit learning (Ellis, 1996) but not implicit learning (Reber et al., 1991; Tagarelli et al., 2011).

There is clearly overlap in these constructs. While there are nuanced differences between these contrasts, for the purposes of the current proposal, I will use the term “Implicit/Procedural Memory” (IPM) to refer to Procedural and Implicit Memory and Learning. In addition, I will use “Explicit/Declarative Memory” (EDM) to refer to Declarative Memory and Explicit Memory and Learning.

Although early work with memory sought to clearly divide multiple forms of memory to more precisely define each construct, the traditionally defined divides between contrasts are not as clear as originally imagined (Dew and Cabeza, 2011). For example, isolating a specific memory process with any given task is nearly impossible. Tasks generally do not break down cleanly along memory distinctions as IPM tasks can also involve EDM and vice versa (Voss and Paller, 2008; Dew and Cabeza, 2011). One issue is that although tasks are frequently described as indices of a particular type of memory (e.g., a “procedural task” or an “explicit memory task”), these tasks rarely index the operation of a single memory system in isolation (Dew and Cabeza, 2011; Voss and Paller, 2008). Additionally, IPM also engages MTL (Schendan et al., 2003; Rose et al., 2002). Schendan et al. (2003) suggest that the mid-MTL (hippocampus and caudate) is involved in sequence learning in both EDM and IPM but engage anterior and posterior regions respectively. Several theoretical frameworks explain MTL involvement in IPM. For example, Shohamy and Turk-Browne (2013) suggest the hippocampus is highly connected to most cortices, including temporal, DLPFC, and the striatum (Suzuki and Amaral, 1994; Goldman-Rakic et al., 1984; Shohamy and Adcock, 2010) and is involved in most behavioral functions. In light of the widespread connectivity of the MTL, Schohamy and Turk-Browne (2013) suggest the hippocampus may exert direct control over the nature of the cognitive representations or modulate cognitive function. In this way MTL is involved in various processing streams. Therefore, the hippocampus is viewed as capable of a wide variety of computations, which are adapted dependent on task demands. Although a number of experimental strategies have been devised to address this issue (e.g., Jacoby, 1991; Schacter, 1987), the lack of a straightforward one-

to-one mapping between memory systems and experimental tasks complicates the interpretation of any experimental finding.

Further, recent evidence suggests that EDM and IPM have some level of interdependence and interactivity and may be differentially engaged due to task demands and the time-course of learning. For example, Poldrack et al. (2001) posit that EDM and IPM networks actually compete during the learning process. Task demands modulate engagement of the MTL (EDM) and striatal networks (IPM) such that tasks promoting EDM showed greater activation in the MTL network whereas tasks promoting IPM showed greater activation in the striatal network. Further, activation of MTL and striatal systems is negatively correlated across participants. In addition, relative reliance on each system can shift over the learning process. For example, Yi et al. (2014) found that early in training, participants used strategies associated with EDM with a gradual shift towards IPM. Interestingly, while these memory processes compete during learning, evidence suggests that the MTL and striatal systems acquire different types of knowledge concurrently (Poldrack et al., 2001). This suggests that if SL is supported by multiple memory systems, the degree to which each memory system is engaged may depend on several factors related to task demands throughout the learning process.

Understanding multiple memory frameworks discussed above in relation to SL can help provide a better understanding of SL as a construct. For example, component processes underlying memory (e.g., EDM, IPM) may also contribute to aspects SL. Furthermore, exploring the interactivity of these previously dichotomized systems can help us better understand the function of SL. For example, one can explore how the

underlying neurobiology may in fact shift during the learning process to rely on different memory systems or have differential activation patterns.

Statistical Learning and Multiple Memory Systems Frameworks

Recent studies have focused on specific connections between SL and multiple memory systems. Many conceptions of SL (see Saffran et al., 1997; Aslin and Newport, 2004; Conway and Christiansen, 2006; Perruchet and Pacton, 2006) assume SL is strictly an IPM process. For example, some theories suggest that SL and IPM measures tap into the same mechanism (Perruchet and Pacton, 2006; Thiessen et al., 2013; 2015; 2017) or that SL is simply a subset of IPM (Conway and Christiansen, 2006). Various lines of research support this assertion. For example, most SL tasks do not give explicit instructions (e.g., Saffran et al. 1997; Aslin and Newport, 2004) and many individuals are not consciously aware of patterns when asked at the end (e.g., Turk-Brown et al. 2009). In addition, SL occurs in infants in the auditory domain, even when infants are distracted with a concurrent drawing task, suggesting SL can occur without direct attention or intention (Saffran et al., 1997). Lastly, SL activates aspects of the frontal (Karuza et al., 2013; Alba and Okanoya, 2008) and striatal (Turk-Browne et al., 2009) systems related to IPM processing.

However, while there is strong evidence linking SL and IPM, recent evidence suggests SL is also supported by EDM. For example, as stated previously, DeKeyser (2003) suggested that EDM relies on conscious awareness and attentional control (executive function), while IPM does not rely on either. Consistent with this, Bertels et al. (2012) found participants have at least some *conscious awareness* of statistical regularities, as tested by confidence ratings during testing, of statistical regularities,

suggesting performance in SL could not be accounted for simply by IPM processing (Bertels et al., 2012). In addition, visual statistical learning (VSL) relies on sustained attention (executive function) such that learning does not occur when participants do not attend consistently to the input stream (e.g., Arciuli and Simpson, 2012). Further, some studies suggest working memory is related to explicit learning (Ellis, 1996), but not implicit learning (Reber et al., 1991; Tagarelli et al., 2011). Shekiela et al. (2016) posited that working memory supports statistical segmentation of structured auditory streams. Furthermore, Shekiela et al. (2016) found that concurrent working memory tasks decreased performance on the SL task irrespective of the modality of the concurrent task (visual, auditory). The authors suggest that SL is therefore related to more domain general processes in working memory. In addition, Yang and Li (2012) found that working memory capacity was correlated with SL performance and this effect was modulated with the type of instructions (explicit or implicit). Similarly, Arciuli (2017) posited an implicit form of working memory may be an additional SL component that affects encoding. Taken together, VSL relies on attention and conscious awareness like EDM and unlike IPM. Understanding the relationship between these constructs (EDM, working memory/executive function, and SL) may provide additional insight into how multiple memory systems support SL.

Turning to brain data, the MTL network has been implicated in SL across modalities (Turk-Browne et al., 2009; Schapiro et al., 2015). This suggests, aspects of SL seem to be supported by multiple memory systems, rather than just IPM. Gomez (2017) examined this possibility through a developmental perspective based in learning and consolidation mechanisms. For example, adults retain statistical patterns learned after a

single exposure, even after a 24-hour period (Kim et al., 2009; Durrant et al., 2012, 2011). However, infants display “fragile” overnight retention up to 15 months (Gomez, 2017; Simon et al., 2016). In addition, infant retention of statistical regularities, unlike in adults, is slow and takes repeated exposure and, crucially, seems to be related to learning processes in the neocortex and striatal networks (Gomez et al., 2017). However, once the hippocampal (MTL) learning system is online, individuals are able to quickly consolidate the statistical information. Furthermore, in adults, hippocampal activity is important for the consolidation of memories overnight (Marshall and Born, 2007). Computational load shifts from hippocampal regions to the neocortex (Davis et al., 2009). However, before two years of age, the necessary connections are not matured and cannot support consolidation (Gomez and Edgin, 2016, Gomez et al., 2017). Before the hippocampal-prefrontal cortex circuit is developed, evidence suggests that SL mainly involves the neocortex and striatal networks with the MTL (hippocampal) network developing more slowly. In conclusion, brain data suggests that the MTL network, a network central to EDM functioning, is activated during SL. Therefore, evidence suggests that brain regions supporting both IPM and EDM are active during SL, contrary to what some theoretical perspectives (e.g., Conway and Christiansen, 2006) that posit SL is strictly an IPM process.

Manipulating SL to Emphasize EDM and IPM Processing

Several studies have manipulated task demands typically associated with the dichotomy between explicit and implicit memory/learning, such as conscious retrieval of information, instruction type, and type of measurement to examine EDM involvement in statistical learning. For example, under some conditions, participants benefit from explicit

instructions (intentional learning condition). The length of item presentation in visual statistical learning (VSL) can affect whether participants benefit from explicit instructions. With slower presentation times, participants presented with explicit instructions have higher learning scores overall (Kachergis et al., 2010; Hamrick and Rebuschat, 2012), but do not with faster presentation times (Arciuli et al., 2014). Arciuli et al. (2014) speculated that their presentation time may have been too fast for participants to benefit from explicit instructions. In addition, in SL with stimuli sequences with simple patterns, participants presented with explicit instructions show more learning (Jimenez et al., 1996), but did not with more complex relationships between items (Frensch and Miner, 1994; Jimenez et al., 1996). Further, performance does not increase with explicit instructions in the auditory modality (Batterink, et al., 2015).

Turning to differences in the developmental trajectory of the influence of EDM and IPM on SL, Yang and Li (2012) found no age-related effects in the implicit instruction condition of an SL task. Older children performed better with explicit instructions suggesting a developmental trajectory regarding influences of EDM and IPM. This is in line with the pattern described in Gomez (2017) in which networks associated with EDM (e.g., MTL learning system) develop more slowly than IPM (e.g., frontal-striatal learning system). This further suggests instruction type differentially interacts with specific task demands related to components of SL. Taken together, pattern complexity, presentation time, modality, and developmental constraints modulate the benefit of explicit instructions and the engagement of EDM in SL.

Even in cases where there is no difference in behavioral data between explicit and implicit instruction conditions, evidence suggests that there are overlapping but

functionally different networks supporting each type of learning related to EDM and IPM (Yang and Li, 2012). Interestingly, Yang and Li (2012) found that both implicit and explicit learners activated similar cortical and subcortical regions associated with the MTL and frontal-striatal networks, but implicit learners showed greater activation of the IFG and caudate than explicit learners. Explicit learners, on the other hand, showed greater activation in the precuneus, typically associated with EDM. These findings are consistent with the assertion presented in Poldrack et al. (2001) regarding interactive networks of activity. Poldrack et al. (2001) posit that EDM and IPM networks actually compete during the learning process. Task demands modulate engagement of the MTL (EDM) and striatal networks (IPM) such that tasks promoting EDM showed greater activation in the MTL network whereas tasks promoting IPM showed greater activation in the striatal network. Further, using connectivity analyses, Yang and Li (2012) found that learning conditions elicit differential patterns of cortical-subcortical connectivity. For example, in the explicit condition, participants recruited more aspects of executive/attentional networks.

Studies found the type of testing (direct, indirect) measures separate contributions from EDM and IPM to SL (Batterink et al., 2015). The authors note that direct measures make reference to studied items and indirect measures examine knowledge through performance-based measures. Therefore, direct and indirect measures should show greater sensitivity to implicit and explicit knowledge. ERP data suggested that performance on the recognition (direct) and target detection task (indirect) were related to EDM and IPM respectively. This suggests that the measure typically used in the tradition of Saffran et al. (1996) is at least in part driven by recall of explicit representation and may underestimate learning as indirect measures seem to measure separate implicit

representations (Batterink et al., 2015). Similarly, grounding theories of SL in multiple memory systems frameworks has important implications for measurement of SL processes and differentiating between online learning processes and the representations created (Siegelman et al., 2017). For example, recent evidence suggests that post learning phase measures in fact tap into representations that may be generated and/or retrieved by differential engagement of EDM/IPM (e.g., conscious/unconscious) (e.g., Batterink et al., 2015; Bertels et al., 2012).

Grounding SL theories in multiple memory systems (EDM/IPM) frameworks/research tradition has interesting implications for the domain general, multimodal regions discussed and may provide additional context for their involvement in SL. For example, SL is subject to age related shifts in engagement of these multiple memory systems (Gomez et al., 2017). These findings are consistent with findings of differential age-related effects of instruction type (Witt et al., 2013) and task dependent effects of engagement of EDM (Arciuli et al., 2014). Taken together, age related and task dependent shifts in engagement of EDM and IPM account for age related and task dependent increases in SL performance (e.g., performance increased as a function of both age and stimulus presentation time) (Arciuli and Simpson, 2011). In addition, integrating EDM and IPM into SL theory suggests there may be differential engagement of multimodal processing regions (MTL, striatal) as a function of task demands.

Integration of SL into more well-established theoretical frameworks allows for questions in SL to be guided by previously discovered phenomena regarding multiple memory systems. For example, SL can be affected by instructions (incidental/intentional divide from explicit and implicit memory literature), measurement (direct/indirect), and

awareness of retrieval (confidence). This further expands investigation of SL to more explicitly examine the nature of the representations generated in SL and to additionally look into online measures of learning (Siegelman et al., 2017). Understanding these additional dimensions and developmental shifts in engagement of multiple memory systems has implications for understanding the componential nature of SL.

Current study

The current study sought to extend the literature by further characterizing the componential nature of SL. In particular, the current study focused on the nuanced relationship between SL and the memory systems discussed above (EDM, IPM). As evidence from both brain and behavior suggests SL seems to be supported by multiple memory systems (dependent on task demands and individual differences) it is important to understand the nature of the relationship between these constructs (i.e., is SL supported by these memory systems directly?). Further, as most behavioral studies have only examined the relationship between SL and multiple memory systems in terms of differences in performance across instructional conditions, this study expanded the scope of the tasks used and further explored the nuanced relationship between these systems by additionally measuring individual differences in EDM, IPM, and executive function. For example, measuring individual differences in EDM and IPM allowed for the comparison to aspects of VSL. If EDM and IPM measures are related or support SL functioning, individual differences in EDM and IPM should correlate with VSL. In addition, if the instructional manipulation affects the underlying processes supporting VSL to differentially emphasize EDM or IPM, the relationship between VSL and multiple memory systems should shift to reflect potential trade-offs in processing (e.g., Poldrack, 2001).

Therefore, in the current study I examined the nature of the relationship between SL and multiple memory systems and whether the manipulating task demands (e.g., instructions, measurement type) on SL shifts the relationship between SL and EDM and IPM. To do this, instructions on a VSL task were manipulated to be explicit or implicit. Then, performance in each condition was compared to examine whether the instructions caused differential engagement of EDM and IPM. In addition, individual performance in each condition was correlated with a battery of EDM and IPM measures to examine the relationship between VSL and EDM and IPM. In addition, I examined whether the instructional manipulation shifted this relationship due to differential engagement of these systems.

In addition, as discussed previously, several cognitive functions are related to both EDM and SL. Inclusion of these measures may further characterize the nuanced nature of the relationship between SL and EDM and IPM. Specifically, I explored the nature of the relationship between SL and executive function/working memory and vocabulary. Further, these measures are more closely associated with EDM than IPM. This allowed for a direct comparison of the relationship between executive function/working memory and EDM. An analysis of the patterning of the relationship between EDM and SL was used to both better characterize how EDM and executive function are related and to provide an additional point of comparison to explore the nuances in the relationship between all of the cognitive functions discussed. Similarly, the relationship between SL and vocabulary was also examined. While representative of individual differences in language processing, vocabulary, for example, may additionally be seen as direct, longer-term semantic (EDM) memory measures. Like with executive function, vocabulary should

be correlated with EDM and pattern similarly with SL. In addition, several studies have suggested that reading and language processes more broadly are related to SL (see Sawi and Rueckl, 2018 for review). Including vocabulary provided preliminary evidence to support this claim. Further, the inclusion of EDM and IPM measures allowed for the clarification of the nuances in the connection between SL and vocabulary.

Overview of Specific Measures

Turning to the specific measures selected, to examine the relationship between SL and multiple memory systems, participants completed one visual statistical learning, three EDM, three IPM, and four cognitive measures (executive function/working memory, vocabulary). The current study used measures/methods established in the literature. Some measures were taken directly from established cognitive batteries. All of these measures were referred to as “established measures.”

Measures were selected on the basis of several factors. First, they needed to have a documented history of use in individual differences analyses with college-aged adults, measuring specific target memory systems or cognitive functions. In addition, for the EDM and IPM measures, reference to specific patterns of brain activity strongly related to the appropriate memory systems was particularly important. Lastly, when available, the psychometric properties of each of the measures (e.g., reliability, distributions) was also taken into account. Specifically, I was interested in examining how well these established measures indexed the given constructs and how reliable each of the measures were. Inconsistencies in the data may have interesting theoretical implications, but they may also be due to noise injected by non-optimal measures. However, while there have been attempts in recent years to develop better measures or improve more established

measures (e.g., Siegelman et al., 2015; Kaufmann et al., 2010), it is important to note there are relatively few in-depth analyses of the psychometric properties of SL and IPM measures. Furthermore, even with these improvements, most tasks have been found to be at least somewhat problematic across the field with regard to reliability and distributional properties of individual differences data (e.g., Siegelman et al., 2015). With this in mind, great care was taken to select the measures that balanced these factors with establishment in the literature with college-aged adults of particular interest.

Visual Statistical Learning (Table 1). The main SL measure was a visual statistical learning (VSL) task based on Siegelman et al. (2017) and Batterink et al. (2015). This paradigm and its variants (cf. Endress and Mehler, 2009; Newport and Aslin, 2004; Siegelman and Frost 2015) are generally seen as the canonical SL measure. The current study used a relatively new self-paced version of the VSL (Siegelman et al. 2017) in order to examine additional aspects of SL not captured by other established methods. The version of the measure that was modified for the current study has been used to examine individual differences in several studies (e.g., Siegelman et al. 2017; Siegelman et al. 2019).

The VSL task was manipulated to emphasize EDM or IPM processing (e.g., Arciuli et al, 2014; Kachergis et al. 2010; Hamrick and Rebuschat, 2012). Two between-subjects conditions were used: the Explicit Learning Condition and the Implicit Learning Condition. SL may be affected by several task demands typically associated with EDM and IPM (e.g., instructions) as discussed above. As the current study used a self-paced version of the VSL (Siegelman et al, 2017), only the instruction type was manipulated. Only manipulating the task instruction type (and not other important factors like pattern

complexity) kept the versions as consistent as possible. Further, while there are well-documented modality-specific components of SL, all tasks used in the current study were visual for consistency and to control for potential interactions with modality.

In addition to the instructional manipulation, both direct and indirect measures of learning in VSL were collected. The direct measure was a basic two-alternative forced-choice task that is most common in the literature. In addition, the indirect measure was a target detection task. These measures are related to EDM and IPM respectively (Batterink et al., 2015). Including both instructional and measurement type manipulations allowed for a more nuanced examination of the relationship between these memory systems. Both the direct and indirect measures of VSL have been used for individual differences analyses in the literature (e.g., Siegelman et al. 2015; Batterink et al. 2015; Otusuka et al., 2016) and the direct measure in particular has been used to examine the relationship between SL and language processing/reading (see Sawi and Rueckl, 2018 for a review).

Explicit/Declarative Memory (Table 2). The EDM measures used were a visual paired associate learning task and a visual object learning task with an immediate and a delayed condition. Each of these measures were taken directly from two well-established cognitive batteries used to examine individual differences and used the same standardized computer programs.

Visual Paired Associate Learning. In visual paired associate tasks, participants w presented with various visual objects and must associate arbitrary visual aspects of these objects such as shape and spatial location. Evidence suggests paired associate learning is dependent on MTL activation (see Krishnan et al. 2016; Suzuki, 2008 for review). Specifically, the Continuous Paired Associate Learning test used in the current

study is part of the Cogstate computerized battery of cognition (Collie, Maruff, Darby, and McStephen, 2003; Pietrzak, Maruff, Mayes, Roman, Sosa, and Snyder, 2008) and is used to measure visual memory with paired associated learning (Maruff et al. 2009). In this specific version, participants associate an abstract shape with a spatial location on the screen. While the Cogstate version has typically been used in older populations, the measure has been normed for college-aged populations and has specific test items and sequences for this group.

Visual Object Learning (Immediate/Delayed). In the visual object learning task, participants are presented with a series of complex shapes. Participants then are presented with more shapes, some new and some old, and decide which ones they have seen before. The visual object learning task and variants (e.g., facial recognition) activate frontal and bilateral anterior MTL regions (e.g., Gur et al., 1997; Jackson and Schacter, 2004) and have been used in functional neuroimaging studies in healthy individuals (e.g., Gur et al., 1997). Specifically, both the immediate and delayed versions of the Visual Object Learning test in the current study are part of the Penn Computerized Neuropsychological Testing computerized battery (Baron et al. 2007) and were used to measure visual learning and memory (e.g., Moore et al. 2015) in college-age adults.

Implicit/Procedural Memory (Table 2). Turning to the IPM measures, each of the three measures used (categorization, artificial grammar learning, serial reaction time task) are considered representative of aspects of IPM. Like the EDM measures, each of these measures were taken or modified from measures established in the literature to be used in individual differences analyses.

Categorization. In the categorization task, participants are shown a series of cards with different symbols on them and make a decision regarding the category these cards represent. Participants are then provided with immediate feedback to guide their future decisions. The categorization task (feedback-based category learning) has been shown to recruit cortical regions related to IPM such as aspects of the frontal-striatal circuit during learning (Poldrack, 2001). The current version of the task was adapted by Marsh et al. (2005) from the seminal (Poldrack, 2001) which found that IPM and EDM have a degree of interactivity and competitively activate during the learning process. The current task measures learning of stimulus-response pairings with immediate feedback and was used in Marsh et al. (2005) to measure individual differences in IPM.

Artificial Grammar Learning (AGL). Artificial grammar learning stems from the research tradition regarding implicit learning, particularly in the extraction of rules from sequences of information and/or the “chunking” of the information based on regularities in the input (see Perruchet and Pacton, 2006 for a review). In artificial grammar learning (AGL), “artificial grammars,” typically created using Markov chains (finite state automata) with particular sets of rules for traversal, generate sequences of stimuli (e.g., letters, symbols, shapes and therefore have a certain set of embedded regularities. Participants are presented with sequences generated by the artificial grammar and then are tested with a grammaticality judgement. Items may be grammatical in two ways: 1) they follow the rules of the artificial grammar and were presented previously (familiarity); 2) they follow the rules of the artificial grammar but were not presented previously (transfer). The version of the task in the current study developed by Pavlidou et al. (2012) only tests grammaticality by looking at previously presented items. The artificial grammar learning

task has been used to examine individual differences in implicit learning in a variety of age groups (e.g., Yang and Li, 2012; Pavlidou et al., 2012). The implicit version of this task has been shown to recruit aspects of the fronto-striatal network (Yang and Li, 2012). In addition, using effective connectivity analysis Yang and Li (2012) found implicit learners display a direct connection between IFG and caudate (cortical-subcortical).

Serial Reaction Time Task (SRT) . The serial reaction time task (SRT) stems from the literature on procedural learning. The SRT is a choice reaction-time task in which participants repeatedly respond to a small set of visual cues, typically by pressing a button paired with each cue. The sequence of cues is structured such that a particular cue is at least somewhat predictable on the basis of the previous cue or series of cues (Nissen and Bullemer, 1987; Robertson, 2007, Siegelman and Frost 2015). Participants were not told to look for any regularities within the input. Over time, typically developing participants become attuned to regularities across stimuli and therefore respond faster to more predictable (structured) items. However, individuals with damage to regions supporting procedural learning either do not show this affect (for a review see Siegert, et al., 2006), or have smaller effects than controls (e.g., Pascual-Leone et al., 1993, Siegert et al., 2006) and thus SRT is strongly related to function in these regions/brain networks. SRT tracks online development of learning and includes involvement of concurrent motor function (Frost et al., 2015). The current version of the task was developed by Siegelman et al. (2015), based on Kaufmann et al. (2010) and has been used to measure individual differences in procedural learning. According to Siegelman et al. (2015), this version has reasonable reliability.

Other Cognitive Measures (Table 2). The cognitive battery included one working memory (WM), two executive function (EF), and one vocabulary measure. These measures were used to explore additional aspects of cognition that are differentially related to EDM and IPM. Like the IPM and EDM measures described above, these measures are taken directly from the literature and have been well-established in individual differences studies. Specifically, the EF/WM measures are standard measures used in computerized batteries of cognition and have been normalized for college-aged populations.

Penn Letter N-Back. The letter-n back is a visual based version of the task that uses progressively more difficult (more taxing on working memory) sequences in which individuals need to hold varying amounts of information in working memory. N-back tasks activate aspects of working memory and attentional networks such as prefrontal cortex and anterior cingulate (e.g., Harvey et al., 2005). The current study used the Penn Letter N-Back task that has been used to measure working memory in the Penn computerized battery (Baron, et al. 2007).

Detection and Identification. In the detection task, participants need to respond as fast as they can when they see a change in the state of a stimuli presented on the computer screen. Detection measures sustained attention and provides a baseline RT to stimuli. In the identification task, participants are given additional rules and inputs for how to respond to the change in the state of the stimuli. Identification also measures sustained attention, but additionally measures aspects of inhibitory control/attentional control due to the more complex decision-making process needed relative to the detection task. Both the detection and identification tasks have been used to measure different aspects of

attention in the Cogstate battery computerized battery and have been normed and used with healthy, college-aged adults (Collie et al., 2008).

Vocabulary. The vocabulary measure from the current experiment was a computerized version of a widely used vocabulary task (Nelson-Denny). Vocabulary provided insight into individual differences in skill with/exposure to various aspects of language (e.g., semantic, orthographic information). Inclusion of Vocabulary provided an additional point of comparison between SL and the memory systems described above. Further, in recent years several lines of research have provided convergent evidence supporting the connection between SL and language (e.g., Arciuli and Simpson, 2012; Frost et al., 2014; Bogaerts et al., 2015). However, an obstacle to fully understanding the theoretical implications of these findings is that the componential nature of SL has not been fully characterized (see Sawi and Rueckl, 2018 for a discussion of this topic). Understanding the relationship between SL, multiple memory systems, and executive function may in turn help better explain how SL and language might interact.

Overview. Individual differences in the IPM, EDM, and cognitive measures (see Table 2) were compared to individual differences in the VSL Explicit and Implicit Learning conditions (see Table 1) to further characterize the relationship between these measures. For example, if the Explicit Learning VSL condition shifts processing to emphasize EDM, performance on each of the Explicit Learning VSL measures should be more strongly related to EDM rather than IPM.

Table 1. Components of Visual Statistical Learning Task

Phase of VSL Task	Aspect of VSL Task	Reference
-------------------	--------------------	-----------

Exposure/Familiarization	Instructional Manipulation	Arciuli et al., 2014
	Cover Task	Arciuli et al., 2012
	Self-paced Task	Siegelman et al., 2015; 2017
Test Phase	VSL-Direct	Siegelman et al., 2015; 2017
	VSL-Indirect	Batterink et al., 2015

Table 2. Measures For Each Construct

Construct	Measure	Reference
Statistical Learning (SL)	Visual Statistical Learning	Siegelman et al., 2017; Arciuli et al., 2012; Arciuli et al 2014; Batterink et al., 2015
Explicit/Declarative (EDM)	Continuous Paired Associates	Cogstate Neurocognitive Battery. Collie et al., 2008
	Visual Object Learning-Immediate	Penn Computerized Neuropsychological Testing Battery. Baron et al. 2007
	Visual Object Learning-Delayed	Penn Computerized Neuropsychological Testing Battery. Baron et al. 2007
Implicit/Procedural (IPM)	Artificial Grammar Learning	Pavlidou et al. 2012 (modified)
	Serial Reaction Time Task	Kaufman et al., 2012; Siegelman et al., 2015
	Categorization	Marsh et al., 2005; 2006
Executive Function (EF)	Identification	Cogstate Neurocognitive Battery. Collie et al., 2008
	Detection	Cogstate Neurocognitive Battery. Collie et al., 2008
Working Memory (WM)	Letter N-Back	Penn Computerized Neuropsychological Testing Battery. Baron et al. 2007
Cognitive/Language	Vocabulary	Nelson, M. J., Brown, J. I., & Denny, M. J., 1960

Specific Aims

The overall aim of the current study was to characterize componential nature of SL by exploring the relationship between SL and the multiple memory systems (e.g., EDM, IPM). The main manipulation in the experiment was the instructions (explicit, implicit) given to participants before the VSL task were manipulated. Learning in the VSL task was also measured using methods differentially related to explicit and implicit learning/memory. Several strategies were used to explore this overall aim. First, as is common in the literature, the effect of instructions on overall VSL performance in both the direct and indirect measures was examined. Several studies have suggested that under some circumstances explicit instructions increased performance (e.g., Kachergis et al., 2010; Hamrick and Rebuschat, 2012). The authors of many of these studies posited that increased performance in the explicit condition suggests increased EDM activity. However, several studies have found that even in the absence of a difference in performance, instructional manipulations cause differential activation of EDM and IPM processes (e.g., increased activation of EDM regions/networks in the explicit condition) (e.g., Yang and Li, 2012).

To further explore the nuances in the relationship between SL and multiple memory systems (that are lost when only examining differences in performance), the relationship between individual differences in VSL performance and performance on several measures of EDM and IPM were explored. Individual differences in a battery of cognitive measures differentially related to EDM and IPM were also examined to further characterize the relationship between VSL and EDM and IPM. All of the measurements used were either taken directly from the literature or were derived from well-established studies as described above.

Aim 1: Examine the effect of instructional condition (explicit, implicit) on Visual Statistical Learning Performance (direct and indirect) and the relationship between each measure.

Aim 2: Examine the relationship between Visual Statistical Learning and Implicit Procedural Memory/Explicit Declarative Memory across instructional conditions (explicit, implicit) and measurement type (direct, indirect).

The measures used in Aims 1 and 2 were well-established in the literature. However, there were several inconsistencies in the relationships between the measures. These inconsistencies may have theoretically interesting implications, but several issues with reliability in inconsistencies in individual differences analyses have been identified. Therefore, in recent years, there has been a push to develop more reliable versions of these measures (e.g., Siegelman et al., 2015; Kaufmann et al., 2010), but issues with reliability remain a concern throughout the literature (Anbal et al., 2019). In Aims 1 and 2, when available, versions of measures from the literature were selected for which in-depth examinations of important psychometric properties were available (e.g., acceptable reliability, distributions). However, while there are legitimate methodological concerns with many of the VSL and IPM measures, these measures have a strong theoretical foundation with well-established neurocognitive evidence. Therefore, they should not be thrown out entirely. Relatively recent advances in analysis and processing techniques (e.g., linear mixed effects modeling, individualized regression) open up the possibility of addressing some of these reliability issues (e.g., statistically controlling for nuisance variables). These advanced analyses provide the opportunity to provide a less noisy version of the data and account for some of the less optimal aspects of the measures. It

is important to note that while these analyses generate cleaner versions of the measures in question, they do not necessarily account for all of the initial inherent theoretical and methodological issues. As such, the current study provided an opportunity to develop new methods for VSL and IPM measures (using advanced statistical analyses) which were potentially more methodologically and theoretically sound. In addition, many measures may be derived from a single task. That is, there were many alternate ways to process the data that may produce new measures related to different processes than the established measures. In Aims 1 and 2, the decision regarding processing and measures was guided by precedent in the literature.

The nature of the results from Aims 1 and 2, the results of the follow-up reliability analysis of the VSL and IPM measures used in Aims 1 and 2, and the methodological and theoretical concerns outlined above led to the development of an additional set of exploratory aims. The main purpose of these exploratory aims was to develop improvements to the measures used in Aims 1 and 2. These improvements allowed for the confirmation of the findings from Aims 1 and 2 and, in some cases, expansion of the understanding of the relationship between VSL and multiple memory systems.

Exploratory Aim 3: Develop improvements to statistical learning and Implicit Procedural Memory measurement to better understand the relationship between VSL and multiple memory systems.

Exploratory Aim 4: Re-examine the findings from Aims 1 and 2 using the updated measures developed in Aim 3 to better understand the relationship between VSL and multiple memory systems.

Study Organization/Presentation Structure

In order to accommodate the complex organization of the current study, each specific aim was treated as a separate experiment. Importantly, each specific aim will contain a methods section with descriptions of the measures most relevant for the given section. In addition, each aim will include a discussion/conclusion to wrap up the important points from the preceding section.

General Method

Participants

Two-hundred and eleven undergraduate students from University of Connecticut participated in Day 1 of the study. However, not all participants completed all of the tasks as some participants opted to leave early. In addition, only 150 participants returned for Day 2. The number of participants in each analysis shifted slightly as not all participants finished all of the tasks in the experiment. No participants were removed as outliers. All participants were neurologically normal. Participants were randomly assigned to one of the two conditions (VSL-Explicit Learning, VSL-Implicit Learning) on Day 1.

Procedure

The experiment took approximately 1 hour and 40 minutes to 2 hours and was completed over two sessions in order to keep participants motivated and ensure the data, particularly on the learning and memory individual difference measures, was useable. Fatigue regarding the potential memory systems used may introduce noise into the experiment. Therefore, it was important to spread out the tasks meant to tap into specific systems as much as possible. In addition, two task orders were used throughout the experiment to help mitigate the effect of task order on the predictor measures.

On the first day, participants first completed the VSL task. In addition, on Day 1, participants completed one EDM measure and one executive function measure. Participants completed the EDM measure last, as having an explicit measure presented first may affect how participants learn in a subsequent VSL measure. On the second day, participants started with the other EDM measure. Then, participants completed two of the IPM measures. Participants then completed part 2 of the EDM measure as there was need for at least a 15-minute delay between the initial learning and delayed post-test. Participants then completed the working memory and vocabulary measure last. There were two presentation orders for the measures in order to mitigate concerns regarding order effects. Presentation order did not have a statistically significant effect on performance.

Aim 1: Effect of Instructional Conditions on SL Performance

Examining the effect of instructions on performance was an important step to examining whether the instructional manipulation had an effect on SL processing. However, it is important to note, differences in performance provide an incomplete picture of the relationship between SL and MMS. For example, even in cases where performance is the same across instructional conditions, the memory systems supporting SL may still be different (Yang and Li, 2012). Therefore, as an initial step to further explore the relationship between SL and MMS, the changes in the relationship between direct (explicit) and indirect (implicit) measures of VSL will also be examined.

Method

Visual Statistical Learning

Exposure Phase

See Table 1 for an overview of the VSL task. The overall structure of the VSL task used was based on the task developed by Siegelman et al. (2017). The VSL task was comprised of two components: an exposure phase and a test phase. Twenty-four abstract shapes were grouped into 8 sets of 3 items (8 triplets) (e.g., Frost et al., 2013; Turk-Browne et al., 2005). In the exposure phase, participants were told they would see a stream of items and would need to pay attention to the order as some of the objects may be repeated.

Cover Task (Arciuli and Simpson, 2012). The cover task was based on Arciuli and Simpson (2012). Each triplet was presented 24 times with 4 of the 24 repetitions showing the first or last shape in the triplet two times in a row. Participants were told to press the “2” key when they saw a repetition to make sure they were paying attention during exposure in both conditions.

Explicit/Implicit Learning Manipulation (Arciuli et al. 2017). The wording of the instructional manipulations was based on Arciuli et al. (2017). In the Implicit Learning condition, participants were told that they needed to watch the order of the shapes for repeats (cover task) and were not told of the test at the end. In contrast, in the Explicit Learning condition, participants were given explicit instructions and cued to the fact that some shapes occurred together and their task would be to find the regularities in the input stream in addition to finding repeated items (cover task). They were also told they would be tested on their knowledge after the exposure phase. The explicit instructions should prompt participants to use explicit learning strategies. The explicit instructions gave participants a specific goal of finding patterns/structure in the input and should prompt

participants to put effort into learning the statistical regularities beyond simply completing the cover task (as described above).

Self-paced Task (VSL-SPT) (*Siegelman et al. 2017*). The self-paced structure of the task was based on Siegelman et al. (2017). Rather than having the items presented at a constant pace as has typically been done in studies using this paradigm, participants advanced the shapes at their own pace by pressing the “1” key (originally developed by Siegelman et al., 2017). RTs for each press were recorded and were the basis of the measure of learning. As participants become sensitive to the statistical regularities in the input stream, participants should be able to use transitional probabilities (TP), or the probability that one item follows another item, to predict the next item in a triplet. Therefore, as each triplet is repeated, RT should decrease for more predictable items (items 2 or 3 in a triplet).

Test Phase

Indirect Measure (VSL-Indirect) (*Batterink et al. 2015*). The Indirect Measure used was based on the structure of Batterink et al. (2015). After exposure, participants completed the target detection task (VSL-Indirect). VSL-Indirect provides additional repetitions of the items and served as both a “post-test” and secondary learning session. In this task, participants were given a target shape and presented with a stream of shapes as presented in the learning phase of the experiment. Participants were asked to press the space bar when they saw the target shape. Each of the 24 abstract shapes (3 shapes create a triplet) were presented as a target 3 times, for a total of 72 trials. The stream of stimuli consisted of 4 of the triplets (12 shapes presented in total) participants experienced during exposure (including the triplet with the target item) in random order.

The triplet including the target was not be allowed to be the first triplet presented. Each triplet was presented an equal number of times throughout the experiment. Similar to the self-paced task, RTs should decrease for more predictable items.

Direct Measure (VSL-Direct) (Siegelman et al., 2015; 2017). The Direct Measure used was based on Siegelman et al. (2015; 2017). After VSL-Indirect, participants completed a direct measure with two types of tasks. Participants were first presented with 64 two-alternative forced-choice (2AFC) trials. Participants saw a target triplet that appeared in the exposure phase ($TP = 1$) and a foil triplet constructed with shapes from exposure but ordered such that the shapes in the triplet never followed one another ($TP = 0$) but occurred in the same position in a separate triplet. Eight foils were constructed in this manner. Individual shapes in a triplet were presented for 500 ms (300 ms ISI), with a 1000 ms gap between target and foil triplets. After the triplets were presented, participants were asked to identify which of the triplets appeared in exposure. Target/Foil pairs were not allowed to repeat. Participants then completed a pattern completion task. Two shapes from a triplet were presented with either the second or third shape missing. Participants were presented with three alternatives to complete the triplet. These alternatives consisted of the correct item, the second shape from an incorrect triplet, and the last shape from another incorrect triplet. All shapes were presented simultaneously. Each triplet was completed two times. The presentation order of the triplets was randomized. These direct measures were collapsed to create a direct measure score (VSL Score). At the end of each component, participants also rated their overall confidence in their answer (1-10) (Batterink et al., 2015) in order to test if participants had explicit knowledge of the regularities.

Results

Learning within VSL Measures

VSL Self-paced Task (VSL-SPT). Learning was measured by examining the difference in RT between item 1 (less predictable) and items 2 and 3 (predictable) in a triplet over the course of the experiment (Siegelman et al., 2017). Incorrect responses were removed from analyses. In addition, RTs faster than 250ms and slower than three standard deviations from individual subject means were windsorized to three standard deviations from the individual subject mean RT. RTs were then log transformed to approximate a normal distribution. Furthermore, cover task items were responded to significantly differently than target items and were removed from further analyses. In addition, the RTs to all of the items in the triplets following a cover task item (e.g., A, A, B, C or C, C, D, E, F) were significantly different than target items not following a cover task item and were therefore also removed from further analyses.

VSL-Indirect. Similar to the VSL-SPT, response time to more predictable items should be faster than less predictable items. Therefore, items at the end of a triplet (item 3) should be responded to faster than items earlier in the triplet (item 1). Incorrect responses were removed from analyses. In addition, RTs faster than 250ms and slower than three standard deviations from individual subject means were windsorized to three standard deviations from the individual subject mean RT. RTs were then log transformed to approximate a normal distribution. This measure provided an indirect (implicit) measure of learning to contrast with the more direct 2AFC and pattern completion trials.

VSL-Direct. Proportion correct for the 2AC and pattern completion trials were combined to create the direct VSL measure.

All three VSL measures (VSL-Direct, VSL-Indirect, and VSL Self-paced) had wide distributions with no floor or ceiling effects. Further, the distributions for each these tasks were all close to normal. Table 3 presents the mean and standard deviations of performance across the three VSL measures (distributions of performance on each of the three measures in appendix).

Table 3. Mean and Standard Deviation of Performance in All Three Basic VSL Measures

VSL Measure	Measurement	Mean	Standard Deviation
VSL Score	Proportion Correct	0.57	0.12
VSL Target Detection	Log RT Difference Score	0.03	0.08
VSL Self-paced Task		0.002	0.03

Note: All measures use basic mean scores as prescribed by the literature.

Performance on two of the three VSL measures indicated participants were able to learn the embedded statistical regularities in the visual input at the group level. Performance on VSL-Direct was significantly better than chance ($t(207)=4747.7$, $p < .001$). In VSL-Indirect the mean difference between items 1 and 3 (unpredictable, predictable) was statistically significant ($t(211)=5.02$, $p < .001$). However, turning to the *VSL Self-paced Task* (VSL-SPT), while the mean difference between items 1 and 2 ($t(206)=2.48$, $p < .001$) was statistically significant, the difference between items 1 and 3 was not ($t(206)=1.6$, $p > .05$) as would be predicted if participants were responding faster to more predictable items. In addition, the differences between items 1 and 2 and 1 and 3 were inconsistent and do not follow a discernable pattern of learning over time (Figure

1) (see Siegelman et al. 2019 for example of consistent learning over time in the VSL-SPT).

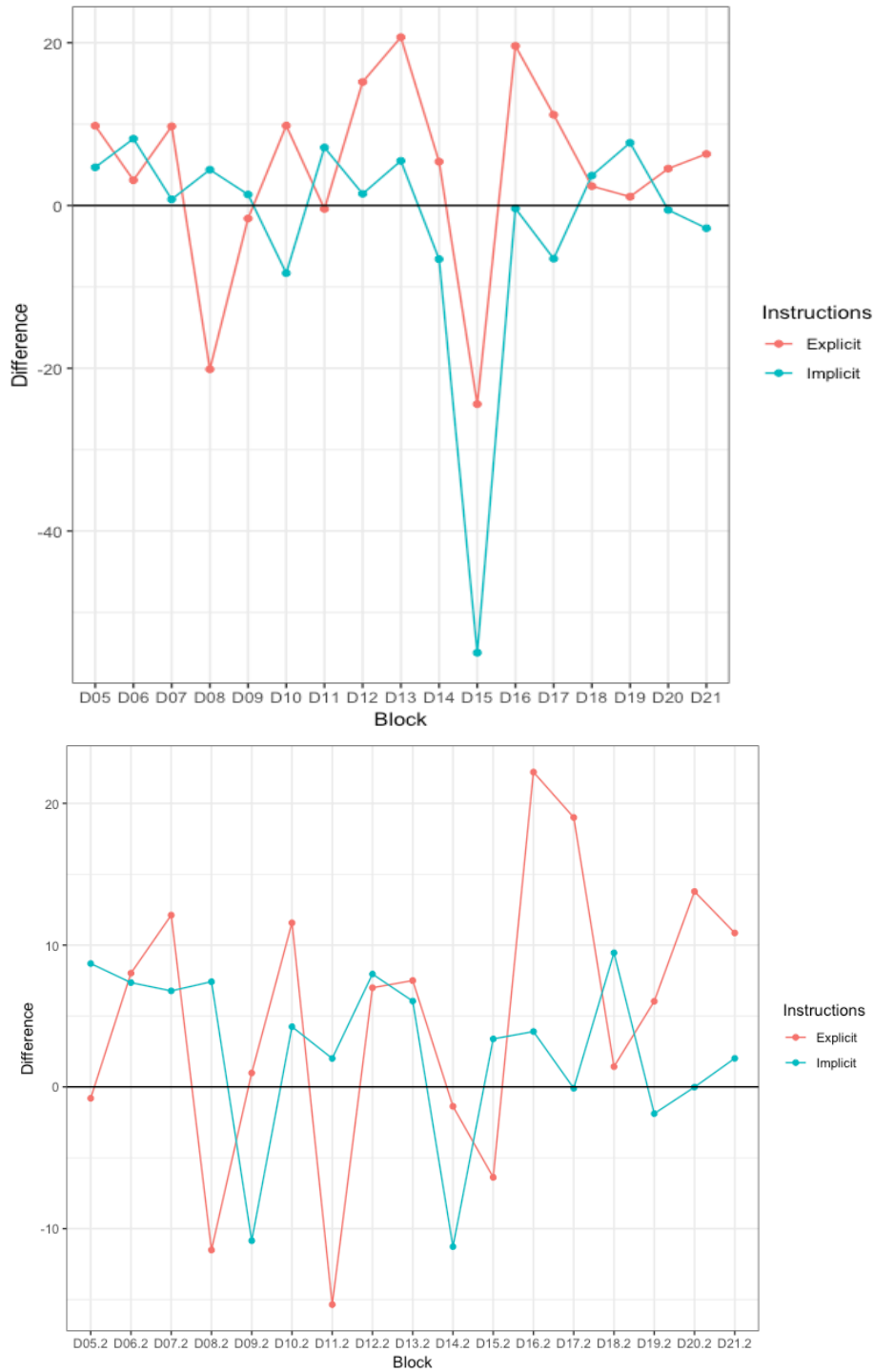


Figure 1. RT Differences between Items 1 and 3 and 1 and 2 (top and bottom respectively). Graphs on the right split participants into Explicit and Implicit instructional conditions. Differences are log transformed.

The lack of consistent differences between items, as found in Siegelman et al. (2019), may be due to the inclusion of the cover task that was run concurrently with VSL-SPT. During the cover task, items 1 and 3 in a triplet were repeated. In addition, participants needed to press a separate button when they saw a repeated item than the button used to move the task forward. As described previously, evidence suggested that the cover task caused a localized disruption to participant RTs, as both cover task (repeated) items and full target triplets directly following cover task items were responded to significantly differently than regular target items. Both cover task items and full triplets following cover task items were removed as described in the Methods. However, it is possible that the cover task caused a more generalized disruption to the learning mechanisms used during the self-paced task. For example, in studies with children, cover tasks are often used (e.g., Arciuli et al., 2012). Some evidence suggests that cover tasks lower overall performance on post-test measures (e.g., 2AFC, pattern completion) (Arciuli et al., 2014). Due to these methodological concerns, the VSL-SPT was removed from further analyses.

Differences in VSL Performance Across Instructional Conditions

Performance on both VSL-Direct and VSL-Indirect did not differ across instructional conditions. In the VSL-Direct, performance in the explicit ($M=.56$) and implicit ($M=.56$) instructional conditions was nearly identical ($t(204.7)=.03$, $p > .05$). Further, an ANOVA was conducted to examine the effect of the instructional manipulation on VSL-Indirect. In VSL-Indirect, neither the overall effect of instructional condition ($F(1, 210) = .387$, $p > .05$) nor interaction between item and instructional condition ($F(2, 420) = .945$,

$p > .05$) were significant. However, the effect of item (predictability) remained significant ($F(2, 420) = .25.78, p < .001$).

Relationship between VSL measures

While the instructional manipulation did not have an effect on performance, it is possible the relationship between VSL-Direct and VSL-Indirect shifted. A shift in the relationship between these measures would suggest that the processes underlying also shifted. This is particularly interesting as VSL-Direct and VSL-Indirect are related to EDM and IPM respectively (Batterink et al., 2015). To examine the relationship between measures, a set of correlations were conducted for all participants, participants in the explicit condition only, and participants in the implicit condition only. To follow-up these results, a series of multiple regression models of increasing complexity were used. First, VSL-Indirect was regressed onto VSL-Direct to confirm the overall correlation. Then instructional condition and the interaction between instructions and VSL-Indirect were added to separate models as predictors. These models were compared to examine whether adding the interaction with instructional condition significantly improved the fit of the model. This process was followed for the following analyses.

Performance on the VSL-Direct and VSL-Indirect, were not significantly correlated ($r(203) = .003, p < .001$). This finding is consistent with Batterink et al. (2015) and may suggest that the VSL-Direct and VSL-Indirect index different aspects of statistical learning. For example, in Batterink et al. (2015), the direct and indirect SL measures from the same exposure task were related to explicit and implicit learning mechanisms respectively.

To examine whether the instructional manipulation shifted the relationship between VSL-Indirect and VSL-Direct, a series of regression models were created and compared. In Model 1.1, VSL-Indirect was regressed onto VSL-Direct. As was the case with the overall correlation, VSL-Indirect was not a significant predictor of VSL-Direct ($b = .01$, $t(202) = .05$, $p > .05$). In Model 1.2, instructional condition was added to the model. Including instructional condition did not improve the fit of the model ($F(1, 203) = .001$, $p > .05$). In Model 1.3, the interaction between VSL-Indirect and instructional condition was added to the model. Including the interaction between VSL-Indirect and instructional condition marginally improved the fit of the model ($F(1, 202) = 3.02$, $p = .08$) over Model 1.2. Furthermore, consistent with the findings from the model comparisons, the interaction term between VSL-Indirect and instructions was marginally significant ($b = -0.37$, $t(202) = -1.74$, $p = .08$) (Fig. 2) and may point to a potential shift in processing such that in the explicit condition as performance on VSL-Direct increased, performance on VSL-Indirect increased. Whereas in the implicit condition, as performance on VSL-Direct increased, performance on VSL-Indirect actually decreased. This suggests that in the explicit condition, performance on VSL-Indirect, a typically IPM-related measure, is more related to performance on VSL-Direct, an EDM measure. However, in the implicit condition, there is a trade-off between VSL-Direct and VSL-Indirect processes. This pattern suggests a trade-off between EDM and IPM processes as found by Poldrack (2001).

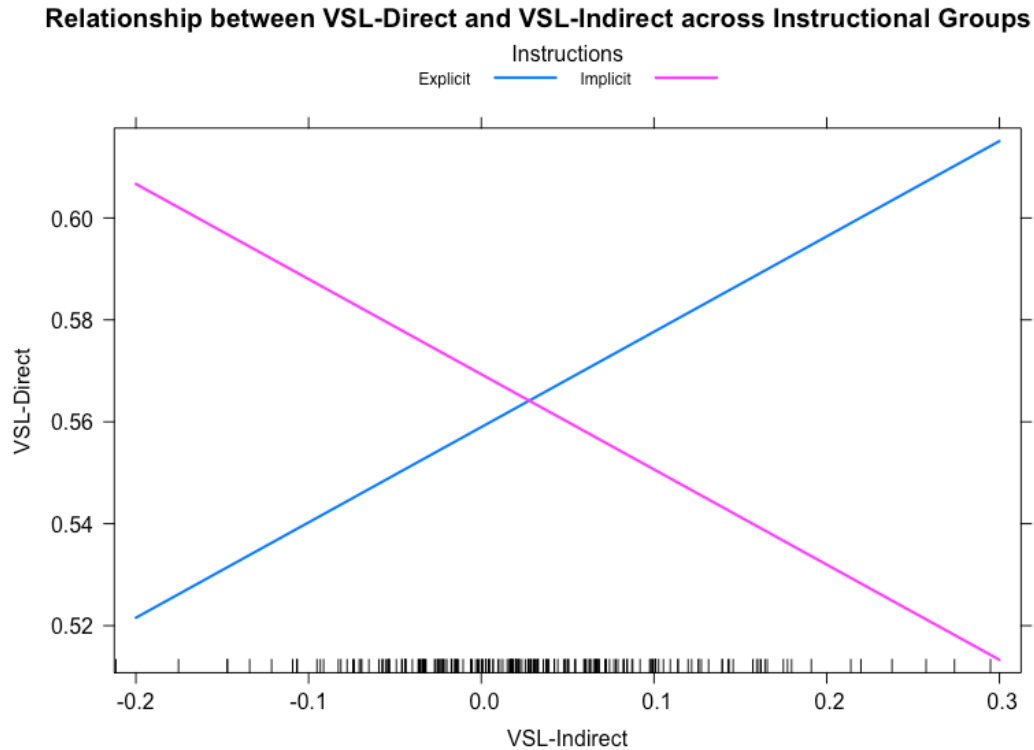


Figure 2. Plot of the marginally significant interaction term (instructional condition) from the regression model. The interaction between VSL-Indirect and instructional condition was regressed onto VSL-Direct.

Aim 1 Interim Conclusion/Discussion

The analyses addressing Aim 1 revealed that: (i) Participants showed learning at the group level for both post-learning measures (VSL-Direct, VSL-Indirect); (ii) the VSL Self-paced Task (VSL-SPT) measure was unusable, most likely due to unintended consequences of the cover task; and (iii) the instructional manipulation did not affect performance on either the VSL-Direct or VSL-Indirect.

It is important to note that a lack of difference in overall performance due to instruction type does not necessarily imply that the underlying processes are the same. For example, Yang and LI (2012) found no differences in performance between explicit and implicit conditions. However, while there were no differences in overall performance, participants in each condition had differential patterns of activation (e.g., greater EDM

activation in the explicit condition). In addition, in implicit and explicit second language learning, Morgan-Short et al. (2012) found no differences in performance. However, ERP results suggests, implicit learners displayed activity more in line with native speakers. Results from Aim 1 indicated there was an interesting marginally significant interaction with instructional condition such that VSL-Direct and VSL-Indirect performance for all participants was positively related in the explicit condition but negatively related in the implicit condition. Therefore, in the explicit condition the processes underlying VSL-Indirect were shifted towards EDM processing resulting in a positive correlation with VSL-Direct. In the implicit condition, however, VSL-Indirect was shifted towards IPM processing and there was a trade-off in performance between VSL-Direct and VSL-Indirect. This is consistent with Poldrack (2001) in which there was a trade-off between EDM and IPM. This trend suggests that, while the overall performance on the VSL tasks was not affected, the underlying processing supporting statistical learning may have shifted (Yang and Li, 2012) and warrant further exploration.

Aim 2: Relationship Between VSL and Multiple Memory Systems Across Instructional Conditions

Aim 2 sought to expand current literature on the link between VSL and multiple memory systems. Most studies examining the link between VSL and multiple memory systems use the difference in performance due to instructional manipulation as the metric for examining whether EDM is involved in SL at all. In Aim 2, I examined the nature of the relationship between SL and multiple memory systems and whether the instructional manipulation shifts the relationship between VSL and EDM and IPM. For example, if

explicit instructions force participants to differentially engage EDM during learning, there should be a stronger relationship to performance on tasks that measure EDM.

While the instructional manipulation did not have an effect on VSL performance, there was a suggestion of a shift in underlying processing in the two conditions. As stated previously, Yang and Li (2012) found that even in the absence of a difference in performance due to the instructional manipulation, functionally different networks related to explicit and implicit memory support SL processing in each condition (e.g., regions related to EDM were activated during the explicit condition). Participants completed three IPM measures, three EDM measures, and four general cognitive measures.

Method

IPM Measures

Implicit learning (artificial grammar learning) (Pavlidou et al. 2012). The current study used the artificial grammar defined in Pavlidou et al. (2012). The procedure from Pavlidou et al. (2012) was modified to include additional repetitions. Twenty-three sequences of increasing length (2-6 abstract shapes) were generated with their grammar. Like with VSL, AGL has an exposure and a test phase. In the exposure phase, participants were trained on the 23 sequences generated by the artificial grammar. The entire sequence was presented simultaneously for 800ms (ISI 200ms). The 23 sequences were presented six times each for a total of 138 sequences. During test there were 64 total trials. Sixteen sequences from the exposure phase were used as “grammatical” items (they were both familiar and followed the artificial grammar). Sixteen “ungrammatical” sequences were generated by changing one item in a grammatical

sequence. Both grammatical and ungrammatical sequences were presented twice. In addition, half of the sequences presented (grammatical and ungrammatical) had high clustering and half had low clustering. Clustering was defined as the total frequency of the bigrams and trigrams that make up a sequence from the exposure phase divided by the number of abstract shapes in a sequence (Pavlidou et al., 2012). For example, in a high clustering sequence, pairs and triplets of abstract shapes within the sequence co-occurred often within the exposure phase across all sequences. The length of high and low clustering items was balanced.

Following the procedure in Pavlidou et al. (2012), during exposure participants were told they would be seeing “alien words” and they needed to pay attention so they could do an activity after. During test participants were told “The alien words you just saw follow some rules. These rules help aliens to put the ‘words’ in the right order. Now you will see more ‘words’ – some follow the same rules, and some do not. You will have to decide which are the ones that follow the same rules.” Then participants were told to press “1” if “you think that a new alien word follows the rules and looks familiar to them or to press “2” if they thought the new alien word does not follow the rules and does not look familiar to them. This grammaticality judgement was used as the basis of learning.

Procedural Learning (Serial Reaction Time Task) (Kaufman et al., 2010; Siegelman et al., 2015). The measure used in the current study was taken directly from Siegelman et al. (2015). In the *serial reaction time task* (SRT), participants saw a shape on the screen in one of four spaces and were asked to press a corresponding button. The order was decided in a probabilistic way (predicting the third item based on the previous two items), with two possible sequence types: sequence A (1–2–1–4–3–2–4–1– 3–4–2–

3) which has an internal probability of 0.85 and sequence B (3–2–3–4–1–2–4–3–1–4–2–1) which has an internal probability of 0.15. The location of the shape can be predicted given the location of the previous two stimuli. For example, if the previous two stimuli were 1 and 4, there was a .85 probability that the next stimulus would be in location 3 (“probable” trial) and a probability of .15 that the item would appear at location 2 (“improbable” trial). The task started with 16 fixed-order practice items. The task included eight trial blocks (120 items per block, 960 total items). The difference in the average response time to the .85 probability items (probable) and .15 probability items (improbable) formed the basis of learning.

Feedback-based category learning (weather prediction task) (Marsh et al., 2005; 2006). The version of the task in the current study was modified from Poldrack et al. (2001) by Marsh et al. (2005; 2006). Participants were presented with a combination of one to four cards in set configurations on a computer screen and were asked whether the cards on the screen represented “sunny” or “rainy”. Participants pressed one of two buttons and received feedback on the response. If the participant did not respond within 1500ms, the trial was recorded as incorrect. After each trial, participants were presented with feedback (e.g., did they answer correctly or what was the correct answer). Each card had a different shape on it (triangle, square, circle, hexagon). There were 14 total combinations of cards presented in specific configurations on the screen and a total of 90 trials. Each card was individually associated with “sunny” or “rainy” with a specific probability. For example, when card 1 (triangle) was presented, 73% of the time the answer was “sunny.” Total percent correct was recorded.

EDM Measures

Visual Paired Associate Learning (*Collie et al., 2008*). The current study used the Cogstate neurocognitive battery. The Continuous Visual Paired Associate Learning task was run directly from the Cogstate software package (*Collie et al., 2008*). In the task, participants were tasked with learning associations between abstract shapes and specific locations on the computer screen. The task occurred in two phases: an initial learning phase and a test phase. In the learning phase, eight randomized abstract shapes with different colors were presented in randomized locations on the screen around a central blue circle. The central blue circle was then replaced with a copy of one of the eight abstract shapes on the screen as a cue to click on the matching shape and associate the location with the abstract shape. Each shape was presented two times during the learning phase. During test, all of the abstract shapes from the learning phase were in the same locations but were covered with blue circles. Participants were then presented with target abstract shapes in the center of the screen as in the learning phase. Participants pressed on the blue circle in the location where the target shape was during the learning phase. There were eight blocks in total. Participants had to answer 30 trials correctly for each block to move onto the next block. Total number of errors was recorded.

Visual Object Learning (*Baron et al., 2007*). The current study used the Penn Computerized Neuropsychological Testing battery (pennncnp.med.upenn.edu) (*Baron et al., 2007*). The Visual Object Learning tasks were presented directly from the Penn Computerized Neuropsychological Testing webpage. The test was presented in two parts. In the first part, participants were shown 10 complex, abstract shapes that they would be tested on later. Each object was presented for 1000ms. In the second part, participants were shown 20 complex, abstract shapes, 10 previously seen and 10 new

(distractor) shapes. Participants did not have a time constraint. For each shape, the participant decided whether they saw the shape in part one on a four-choice scale (“definitely not,” “probably not,” “probably yes,” or “definitely yes”). After 15-20 minutes, the participant completed part 2 of the experiment again (delayed condition). Participants completed IPM or EF/WM measures between the immediate and delayed condition.

Cognitive Measures

Letter N-Back, Working Memory (WM) (*Baron et al., 2007*). The current study used the Penn Computerized Neuropsychological Testing battery (pennncnp.med.upenn.edu) (*Baron et al., 2007*). The Letter N-Back was presented directly from the Penn Computerized Neuropsychological Testing webpage. In this task participants watched a series of letters that are shown on the screen one at a time. Participants pressed the space bar according to three types of rules (0-back, 1-back, and 2-back). During 0-back, participants pressed space when they saw an X. In the 1-back condition, participants pressed space when the letter was the same as the previous letter. Lastly, in the 2-back condition, participants pressed the spacebar when the letter was the same as the letter two letters back. Mean RT for across all trials was used as the measure of WM.

Detection, Executive Function (EF) (*Collie et al., 2008*). The current study used the Cogstate neurocognitive battery. The Detection task was run directly from the Cogstate software package (*Collie et al., 2008*). During the Detection task, playing cards were presented one at a time on a computer screen. Each playing card was initially

presented face-down. Participants were told to left-click on the mouse as fast as possible when the card flipped face up. Participants had 1000ms to respond. Participants received different auditory feedback for correct responses and if they responded too quickly or failed to respond. Participants continued responding until 30 correct responses were recorded.

Identification, Executive Function (EF) (*Collie et al., 2008*). The current study used the Cogstate neurocognitive battery. The Identification task was run directly from the Cogstate software package right after the Detection task was completed (*Collie et al., 2008*). During the Identification task, like with the Detection task, playing cards were presented one at a time on a computer screen. Each playing card was initially presented face down. Unlike the Detection task, the cards were either red or black. When the card flipped face up, participants were told to left-click on the mouse as fast as possible if the card was black or right-click if the card was red. Participants had 1000ms to respond. Participants received different auditory feedback for correct and incorrect responses. Responses were considered incorrect if the wrong click was used or if participants responded too slowly or too quickly.

Nelson-Denny Vocabulary Test (*Nelson et al., 1960*). The Vocabulary task used in the current study was a modified, computerized version of the task used by Nelson et al. (1960). Participants were given unlimited time to complete a 50-question vocabulary test. Each question was in the form of a sentence with a missing word (e.g., to be intelligent is to be ____). Participants were given four options to complete the sentence. Questions were of increasing difficulty.

Results

Once the basic information regarding performance on the IPM, EDM, and Cognitive measures (see Table 2 for overview) was examined (e.g., distributions, correlations within and between the constructs), hierarchical regression was used to explore the relationship between these constructs and VSL. Specifically, for each measure a series of regression models with increasing complexity (instructional manipulation, interaction with instructional manipulation) were created. These models were then compared to examine whether the inclusion of these additional predictors, such as the interaction with the instructional manipulation, improved the fit of the model using analysis of variance. If the interaction term with the instructional condition improved the fit of the model, then there was evidence that the type of instructions shifted the underlying processes supporting VSL.

Table 4. Correlations Between IPM, EDM, and Cognitive Measures

		Implicit/Procedural (IPM)			Explicit/Declarative (EDM)			Cognitive Measures		
		AGL Prop	AGL d'	SRT	VOL-I	VOL-D	CPAL	ID	DET	L-NB
IPM	AGL Prop Correct									
	AGL d'	0.59**								
	SRT	0.10	0.03							
EDM	Visual Object Learning-Immediate	0.04	-0.04	-0.07						
	Visual Object Learning-Delayed	0.06	-0.01	0.00	0.57**					

	Continuous Paired Associates	-0.02	0.05	0.10	0.13	0.06				
Cognitive Measures	Identification	-0.04	-0.09	-0.23	0.04	-0.04	-0.10			
	Detection	0.00	-0.05	-0.13	-0.04	-0.04	-0.05	0.58**		
	Letter N-Back	-0.03	0.00	-0.11	0.01	0.00	-0.05	0.40**	0.35**	
	Vocabulary	0.03	-0.01	-0.09	0.27**	0.25**	-0.03	-0.11	-0.11	-0.09

Note: AGL Proportion Correct (AGL Prop), AGL dPrime (AGL d'), Serial Reaction Time (SRT), Visual Object Learning-Immediate (VOL-I), Visual Object Learning-Delayed (VOL-D), Identification (ID), Detection (DET), Letter N-Back (L-NB)

IPM Measures

IPM Analyses/Measurement

Artificial Grammar Learning (AGL). Learning in the artificial grammar task was measured in two ways. First, total proportion correct for the 64 grammaticality judgements was used to measure implicit learning of the sequences. In addition to the total proportion correct, dPrime was also used. dPrime is derived from signal detection theory and reflects the ability of a participant to accurately discriminate between stimuli. dPrime is calculated by subtracting the z-scored proportion of false positive responses from the z-scored proportion of true negative responses. This is particularly useful when participants are presented with one image and asked if they have seen the image before. dPrime accounts for response bias and helps provide a clearer measurement.

Serial Reaction Time Task (SRT). The RT difference between probable (.85) and improbable (.15) responses from blocks 3 to 8 was the dependent measure as described in Kaufman et al. (2010). Incorrect trials were removed from analyses. In addition, RTs faster than 150ms were removed from analyses and RTs slower than three standard deviations from individual subject means were windsorized to three standard deviations

from the individual subject mean RT. RTs were then log transformed to approximate a normal distribution.

Categorization. Proportion of correct responses was the main dependent measure. An adjusted score was also explored. Using the raw correct/incorrect values may be misleading as cards are probabilistically associated with “sunny” or “rainy” (e.g., card 1 has a 73% chance of meaning “sunny”). If participants become attuned to the probabilistic relationships between the categories throughout the task, when participants see card 1, for example, they should be more likely say the card is “sunny” (this would be incorrect 27% of the time). The adjusted measure accounted for this issue in scoring. However, accounting for this did not have a significant effect on overall group-level performance and was not examined further.

Individual Differences in IPM Measures

All three IPM measures (AGL, SRT, and Categorization) had wide distributions with no floor or ceiling effects. Furthermore, the distributions for each these tasks were all close to normal (appendix). Performance on two of the three AGL measures indicated participants were able to learn at the group level. Performance on the AGL was significantly better than chance ($t(169)=11.89, p < .001$). In the SRT the mean difference between unpredictable and predictable items was statistically significant ($t(163)= -13.15, p < .001$). However, turning to the Categorization task, performance was not significantly above chance at the group level ($t(169)=1.55, p > .05$) and will therefore not be included in further analyses.

Interestingly, while the AGL measures (proportion correct, dPrime) were significantly correlated with each other ($r = .59$, $p < .001$), they were not correlated with the SRT measure ($r = .098$, $r = .03$) (see Table 4).

EDM Measures

EDM Analyses/Measurement

Continuous Paired Associate Learning. Total number of incorrect responses over the 8 blocks of the experiment was the dependent measure. The distribution was skewed such that participants had relatively few incorrect responses. A log transformation was used to approximate a normal distribution. In addition, scores were inverted for ease of interpretation.

Short Visual Object Learning-Immediate, Delayed. Both measures used proportion correct.

Individual Differences in EDM Measures

All three EDM measures had wide distributions with no floor or ceiling effects. Furthermore, the distributions for each these tasks were all close to normal (appendix). While performance on the Visual Object Learning-immediate and -delayed conditions were significantly correlated ($r = .57$, $p < .001$), they were not correlated with the other EDM measure (Continuous Paired Associate Learning) (see Table 4).

Other Cognitive Measures

The outcome measures Detection and Identification (executive function) were the log-transformed mean RT across 30 total correct responses. The Letter N-back (working memory) used the log-transformed *median* RT across the three types of tasks. Vocabulary used total proportion correct.

All four cognitive measures had wide distributions with no floor or ceiling effects. Furthermore, the distributions for each of these tasks were all close to normal (appendix). All executive function/working measures were correlated with each other, but Vocabulary was not correlated with any of the other cognitive measures (see Table 4).

Relationship between IPM, EDM, and Cognitive measures

There were no significant correlations between IPM and EDM measures (see Table 4). In addition, there were no significant correlations between EDM and executive function measures as would be predicted from DeKeyser (2003). However, there was a significant negative correlation between SRT (IPM) and Identification ($r = -.23$, $p < .001$), such that individuals with faster RTs on the identification task had larger predictability effects in the SRT task. Turning to Vocabulary, there were no significant correlations between Vocabulary and IPM or executive function. However, vocabulary was significantly correlated with both visual object learning measures, such that individuals with higher visual object learning (immediate, delayed) performance had higher vocabulary performance.

VSL and IPM

To explore the relationship between VSL and IPM, hierarchical regression analyses were conducted. A series of regression models of increasing complexity were created (e.g., including instructional condition, interaction with instructional condition). These models were then compared to determine whether there was a difference in the relationship between VSL and multiple memory systems across instructional conditions. First, Model 2 examined the relationship between SRT and VSL-Direct. In Model 2.1, SRT was regressed onto VSL-Direct. Then the instructional manipulation was added to the

model in Model 2.2 and the interaction between instructional condition and SRT were added to the regression model in Model 2.3. These models were compared in a step-wise manner to examine whether adding the interaction with instructions significantly improved the fit of the model. This process was followed for the following analyses.

In Model 2.1, SRT was a significant predictor of VSL-Direct ($b = .63$, $t(146) = 1.95$, $p < .05$). In Model 2.2, instructional condition did not improve the fit of the model ($F(1, 145) = .046$, $p > .05$). In addition, in Model 2.3, the interaction between SRT and instructional condition did not improve the fit of the model over Model 2.2 ($F(1, 145) = 1.28$, $p > .05$). Importantly, in Model 2.3, the interaction term between SRT and instructions was not significant ($b = .75$, $t(145) = 1.34$, $p > .05$).

Model 3 explored the relationship between AGL and VSL-Direct. Models 3.1-3.3 followed the same structure as Models 2.1-2.3. In Model 3.1, AGL was not a significant predictor of VSL-Direct overall ($b = .13$, $t(153) = 1.95$, $p > .05$). Neither instructional condition (Model 3.2) ($F(1, 153) = .065$, $p > .05$) nor the interaction between AGL and instructional condition (Model 3.3) ($F(1, 153) = 1.36$, $p > .05$) improved the fit of the model. In addition, in Model 3.3, the interaction term was not significant ($b = 0.31$, $t(151) = 1.16$, $p > .05$).

None of the IPM measures predicted VSL-Indirect performance and there were no significant interactions with instructions.

VSL and EDM

As was the case with the IPM measures, a series of regression analyses were conducted. The same model generation and comparison techniques were used. Model 4 looked at the relationship between paired associate learning and VSL-Direct. In Model

4.1, paired associate learning was a significant predictor of VSL-Direct ($b = 0.03$, $t(201) = 4.02$, $p < .05$). Model 4.2 added the instructional manipulation as a predictor to the model. Instructional condition did not improve the fit of the model ($F(1, 200) = .06$, $p > .05$). Model 4.3 added the interaction between instructional condition and paired associate learning. The interaction term did not improve the fit of the model ($F(1, 199) = .30$, $p > .05$). However, in Model 4.3, while there was a significant relationship between VSL-Direct and paired associate learning $b = 0.02$, $t(199) = 2.38$, $p < .001$), the interaction between paired associate learning and instructions was not significant ($b = 0.007$, $t(199) = .51$, $p > .05$). This suggests that the instructional condition did not shift the relationship between VSL-Direct and paired associate learning.

Model 5 examined the relationship between VSL-Indirect and visual object learning-immediate. In Model 5.1, visual object learning-immediate was not a significant predictor of VSL-Indirect ($b = -0.06$, $t(149) = -.83$, $p > .05$). However, adding instructional condition to the model significantly improved the fit of the model (Model 5.2) ($F(1, 149) = 4.01$, $p < .05$). The interaction between instructional condition and visual object learning-immediate marginally improved the fit of the model (Model 5.3) ($F(1, 148) = 3.78$, $p = .053$) suggesting there may be an interaction. This implies that the instructional condition caused a shift in the processes underlying VSL-Indirect such that in the explicit condition VSL-Indirect was supported more by EDM. However, in the implicit condition, there was a trade-off between VSL-Indirect performance and EDM suggesting a shift towards IPM (Poldrack, 2001; trade-off between EDM and IPM processes). As the model comparisons suggest, in Model 5.3, both instructional condition ($b = 0.20$, $t(147) = 1.70$, $p = .08$) and

the interaction between visual object learning-immediate and instructional condition ($b = 0.27$, $t(147) = -1.84$, $p = .053$) were marginally significant (Fig. 3).

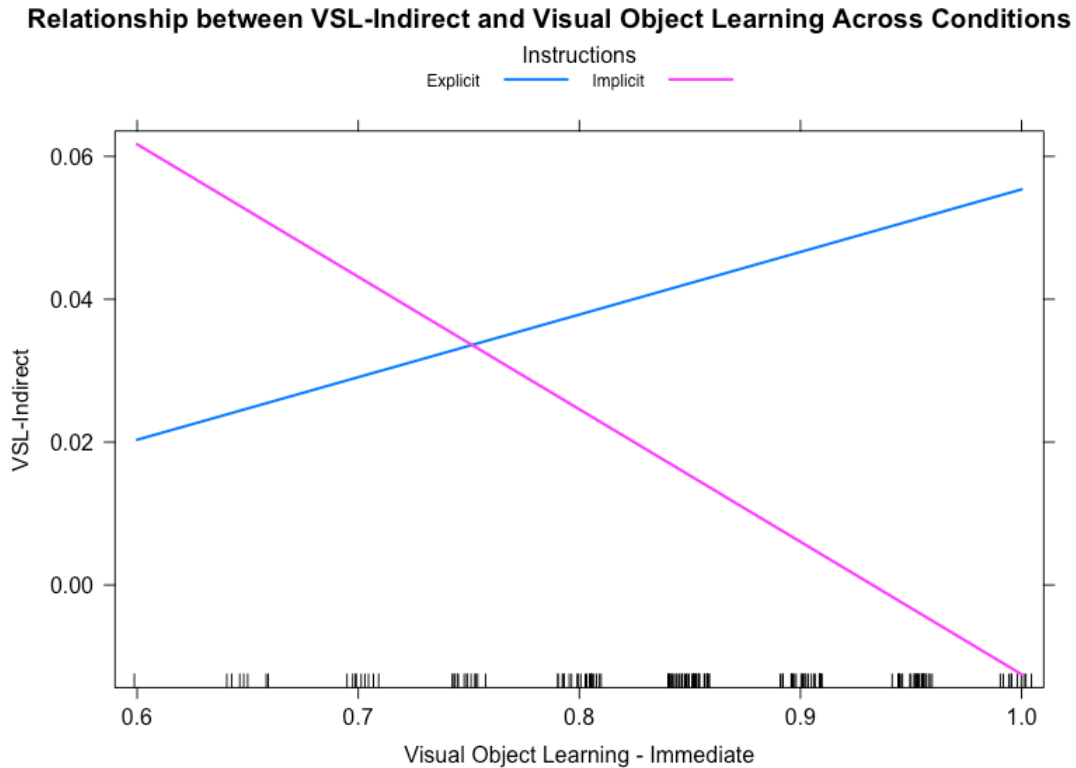


Figure 3. Plot of the interaction term (instructional condition) from the regression model. The interaction between CPAL and instructional condition was regressed onto VSL-Direct. There was a strong relationship between variables, but it was not affected by instructional condition.

VSL and Cognitive Measures

Turning to the relationship between VSL and the cognitive measures described previously (executive function, working memory, vocabulary), hierarchical regression was used as with EDM and IPM. First, in Model 6, the relationship between Identification (executive function) and VSL-Direct was examined. In Model 6.1, Identification was regressed onto VSL-Direct. Identification was a significant predictor of VSL-Direct ($b = -0.27$, $t(204) = -2.22$, $p < .05$). In Model 6.2, instructional condition was added to Model

6.1. Instructional condition did not improve the fit of the model ($F(1, 202) = .05, p > .05$). In Model 6.3, the interaction between Identification and the instructional manipulation was added to the model. The interaction did not improve the fit of the model ($F(1, 202) = .32, p > .05$). Consistent with these results, in Model 6.3, the interaction between Identification and instructional condition was also not significant ($b = 0.14, t(204) = .57, p > .05$).

Similarly, Model 7 examined the relationship between Detection (executive function) and VSL. In Model 7.1, Detection was regressed onto VSL-Indirect. Detection was a not significant predictor of VSL-Indirect ($b = -0.27, t(204) = -2.22, p < .05$). Including instructional condition (Model 7.2) as a predictor did not improve the fit of the model ($F(1, 202) = 2.09, p > .05$). However, adding the interaction between Detection and instructional condition (Model 7.3) significantly improved the fit of the model ($F(1, 201) = 5.89, p < .05$). Furthermore, in Model 7.3, the interaction term (Detection and instructional condition) was significant ($b = 0.28, t(201) = 2.46, p < .05$) (Fig. 4) such that in the explicit condition, performance on the Detection task increased, performance on VSL-Indirect also increased. However, in the implicit condition, performance on the Detection task increased, VSL-Indirect performance actually decreased. This suggests that in the explicit condition executive function supports learning in VSL-Indirect but in the implicit condition there is a trade-off between VSL-Indirect and EDM (Poldrack, 2001).

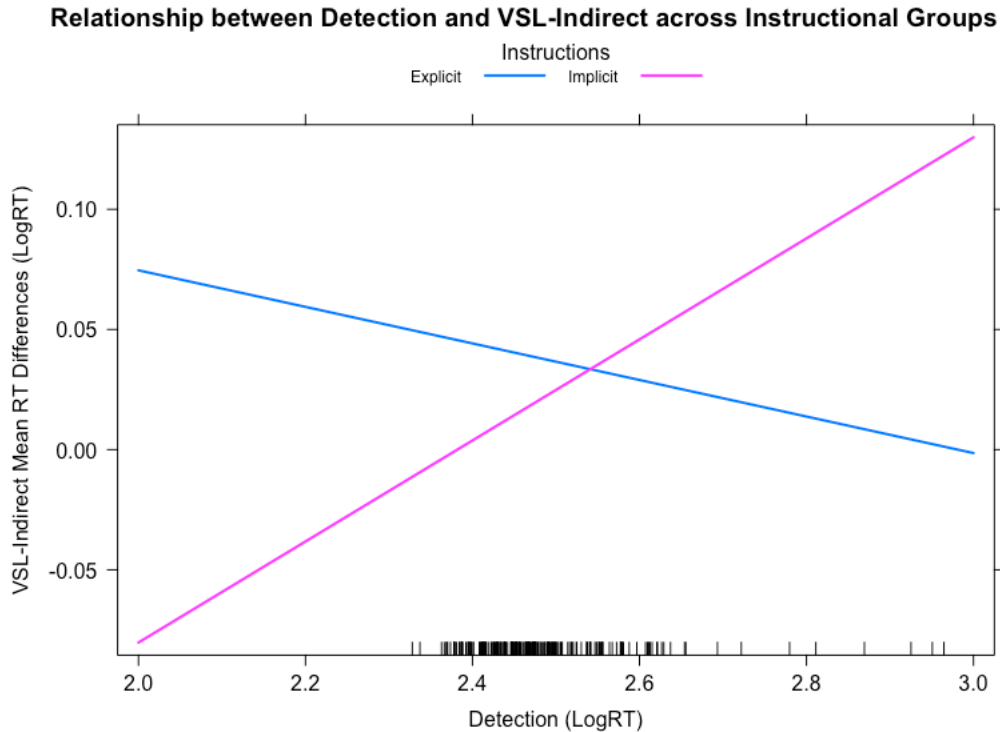


Figure 4. Plot of the interaction term (instructional condition) from the regression model. The interaction between Detection and instructional condition was regressed onto VSL-Indirect.

In Model 8, the relationship between VSL and Vocabulary was examined. In Model 8.1, Vocabulary was a significant predictor of VSL-Direct ($b = 0.18$, $t(142) = .39$, $p < .001$). Neither instructional condition (Model 8.2) ($F(1, 141) = .18$, $p > .05$) nor the interaction between Vocabulary and instructional condition (Model 8.3) ($F(1, 140) = .16$, $p > .05$) improved the fit of the model. In addition, in Model 8.3, the interaction between Vocabulary and instructional condition ($b = 0.05$, $t(140) = .39$, $p > .05$) was not significant.

Both VSL-Direct and Visual Object Learning-Immediate/Visual Object Learning-Delayed (EDM measures) were related to Vocabulary. Paired Associate Learning (a separate EDM measure) also predicted VSL-Direct. Therefore, a multiple regression was used to explore how VSL-Direct, EDM, and Vocabulary were related. VSL-Direct, Visual Object Learning-Immediate and Visual Object Learning-Delayed were regressed onto Vocabulary. Interestingly, even when controlling for both Visual Object Learning

measures, VSL-Direct was a significant predictor of Vocabulary ($b = 0.2$, $t(132) = 2.20$, $p < .05$).

Aim 2 Interim Conclusion/Discussion

In Aim 2, I examined the relationship between VSL and IPM, EDM, and a battery of cognitive measures across explicit and implicit instructional conditions and measurement types. Aim 2 provided insight into the relationship between these constructs and expanded on previous findings. Several interesting findings from this aim will be highlighted below.

1. Individual performance in both EDM and IPM predict VSL-Direct.

Performance on continuous paired associate learning (EDM) and SRT (IPM) predicted performance on VSL-Direct (explicit). This relationship was not affected by instructional condition. This is particularly interesting as many conceptions of SL assume SL is an IPM process or simply a subset of IPM (see Saffran et al., 1997; Aslin and Newport, 2004; Conway and Christiansen, 2006; Perruchet and Pacton, 2006). However, as stated above, recent evidence from brain and behavior suggests EDM involvement in SL (e.g., Hamrick & Rebuschat, 2012; Turk-Browne et al., 2009; Schapiro et al., 2015). The current findings extend the literature as most previous explorations of the involvement of EDM in SL used an instructional manipulation (e.g., Hamrick and Rebuschat, 2012, Frensch & Miner, 1994; Jimenez et al., 1996) and could only infer EDM involvement due to differences in performance across instructional conditions. The involvement of both EDM and IPM in VSL is consistent with the brain data suggesting

activation in both MTL (EDM) and frontal-striatal (IPM) circuits across modalities (see Sawi and Rueckl, 2018; Frost et al., 2015 for a review).

2. Instructional condition may cause differential engagement of EDM and IPM during VSL-Indirect (marginally significant).

Unlike with VSL-Direct, the relationship between visual object learning-immediate (EDM) and VSL-Indirect was marginally affected by the instructional manipulation. In the explicit condition, there was a slight positive relationship such that as performance on the visual object learning task-immediate increased, performance on VSL-Indirect also increased. However, in the implicit condition, there was actually a negative relationship such that individuals with higher visual object learning-immediate scores performed worse on VSL-Indirect. These results provide additional evidence that VSL is related to EDM. However, VSL-Indirect, a more implicit measure of VSL (Batterink et al., 2015), may differentially engage EDM due to instructional manipulation (correlation or trade-off).

3. Different aspects of executive function were related to VSL-Direct and VSL-Indirect, but the pattern established with EDM across instruction types (no shift in VSL-Direct, shift in VSL-Indirect) remained.

Consistent with several previous findings (Arciuli et al., 2014), performance on VSL-Direct was predicted by identification (sustained attention, inhibitory control). VSL-Indirect performance was also predicted by executive function performance but with the

version of the task that relates more to sustained attention (detection) rather than the ability to apply specific instructions and select the proper response (inhibitory control). In addition, whereas the relationship between executive function and VSL-Direct did not shift due to instructional condition, the relationship between VSL-Indirect and detection was shifted due to the instructional manipulation. In the explicit condition, better detection skill resulted in larger VSL-Indirect difference scores as one might expect if the explicit instructions shift processing towards EDM. In the implicit condition, there was a strong negative relationship, such that better Detection skill actually resulted in significantly smaller effects almost to the point where unpredictable items were responded to faster than predictable items. The interaction between detection and instructional condition is consistent with prediction that EF and EDM are related processes. Furthermore, this result mirrored the connection between working memory and SL found in Yang and Li (2012). Controlling for age, in the explicit condition, working memory strongly predicted SL, but in the implicit condition this effect went away entirely.

4. Across the board, the instructional condition did not have an effect on the processes underlying VSL-Direct (Explicit), but VSL-Indirect (Implicit) was affected by the instructional condition.

Interestingly, whereas the instructional condition did not shift the relationship between VSL-Direct and paired associate learning and Identification (EDM and EF respectively), instructional condition did have an effect on the relationship between VSL-Indirect and Detection. This pattern was present across the board for VSL-Direct

(continuous paired associate learning, SRT, identification, vocabulary) and VSL-Indirect (visual object learning-immediate, detection).

This may be due to the fact as a direct measure, VSL-Direct may already be shifted towards explicit processing, and implicit instructions do not shift the underlying processes enough to see tangible differences in the relationship between VSL-Direct and EDM. Conversely, as an indirect measure, VSL-Indirect may be shifted towards implicit processing. However, unlike VSL-Direct, VSL-Indirect is more subject to differential processing as, VSL-Indirect may be more like a learning measure than the completely post-learning VSL-Direct. For example, before VSL-Indirect, participants in the explicit instructions condition were reminded that they were looking for specific patterns. After seeing the initial target item, participants saw each of the complete triplets, one item at a time for an additional 36 times. In addition, participants saw each item for 500ms. This was within the range in which Arciuli et al. (2014) speculated participants could engage explicit memory due to explicit instructions. These factors may have caused a shift in processes underlying VSL-Indirect not present in VSL-Direct.

5. VSL-Direct performance predicted performance on the vocabulary task, controlling for EDM.

Interestingly, both VSL-Direct and visual object learning-immediate/delayed (EDM) were correlated with vocabulary. The similarity in patterning may additionally support the conclusion that VSL-Direct and EDM are related constructs. The multiple regression including VSL-Direct and both visual object learning measures as predictors of vocabulary

helped clarify the complex relationship between these constructs. It is unsurprising that VSL-Direct was related to EDM as VSL-Direct has been established as a more explicit measure of VSL. However, even controlling for EDM, VSL-Direct was still a strong predictor of vocabulary performance. Several frameworks have suggested that the MTL plays a critical role in the learning of arbitrary associations like may be measured by vocabulary (McClelland et al., 1995; Squire, 1992). This suggests that, like with paired associate learning, MTL mediation would be particularly important in learning semantic information. Consistent with this assertion, vocabulary seems to be related to EDM through the more explicit aspects of VSL (VSL-Direct). In addition, VSL-Direct may be related to vocabulary through MTL mediation and the abstraction of arbitrary associations (McClelland, McNaughton, and O'Reilly, 1995; Squire, 1992) as is important for learning semantic information.

6. While performance on IPM and EDM measures were in an acceptable range, only performance on measures derived from the same task were correlated.

Participants showed learning across all of the EDM and IPM measures except for the Categorization task. Interestingly, none of the EDM measures or IPM measures correlated with measures described under the same construct of memory. For example, while the visual object learning measures were related as they used the same materials delayed by 15-20 minutes, these measures did not correlate with the paired associate learning task. This may suggest that these measures actually index separate constructs.

This would be in line with the speculation from Gur et al. (2009) that suggested the nature of the paired associate learning task uses more working memory and executive function/attentional resources than a task like the visual object learning task and is less of a “pure” measure of explicit memory. However, the paired associate learning task is also not correlated with any of the executive function measures.

Similarly, IPM measures such as the SRT and the AGL were not correlated, again putting into question the idea that these measures may be separate indexes of an overall “IPM” construct. In addition, unlike as predicted, EF/attention was actually correlated with SRT and not any of the EDM measures.

Experiment Part 1 (Aim 1 and 2) General Discussion

Overall, based on the measures established in the literature (e.g., Arciuli et al., 2012; Siegelman et al., 2015; Batterink et al., 2015), performance on VSL was not affected by instructional condition (Aim 1). In Aim 1 there was a hint of a shift in the relationship between VSL-Direct and VSL-Indirect. However, the interaction with the instructional conditional was only marginally significant. This pattern may either be driven or occluded by measurement issues inherent to using proportion correct and mean difference scores (e.g., noisy signal) as typically prescribed by the literature. Improvements to these measures may help confirm or disconfirm this pattern of results.

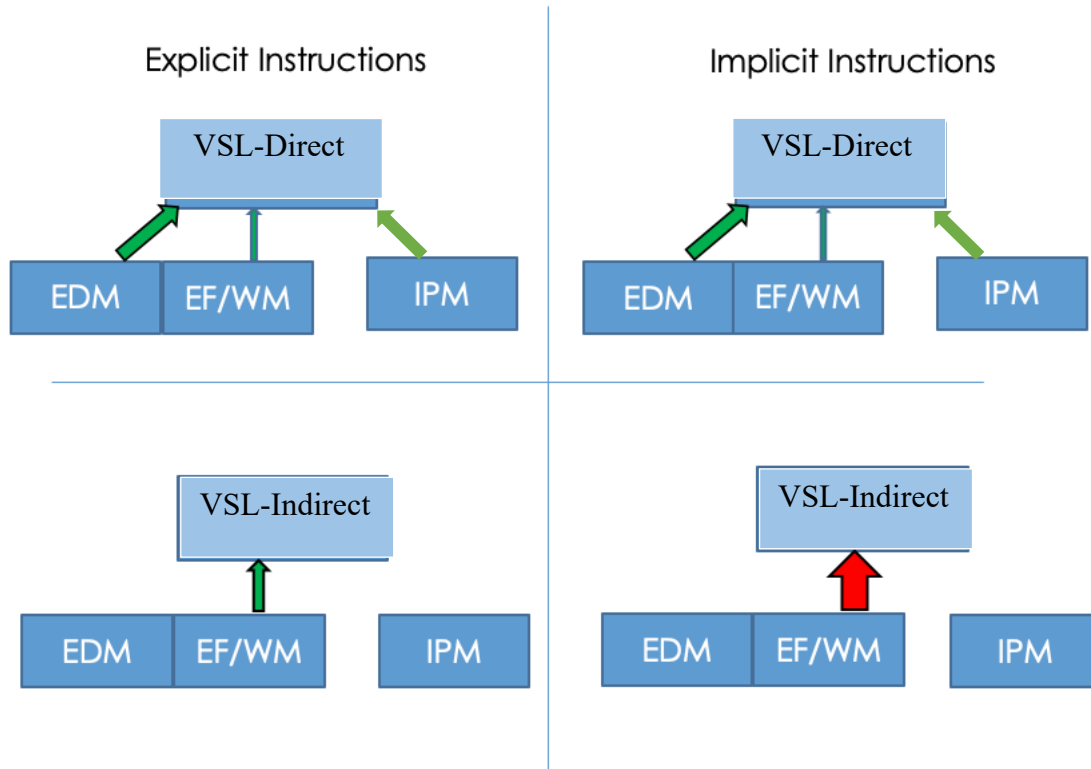


Figure 5. Relationship between VSL and EDM, EF/WM, and IPM. Explicit instructions on the left, Implicit instructions on the right. Direct (explicit) measures on the top row, indirect (implicit) measures on the bottom row. Green arrows represent a positive relationship, red arrows represent a negative relationship. Arrow width represents the strength of the connection.

Aim 2 sought to explore the relationship between SL and multiple memory systems that are missed when only looking at the effect on performance (see Fig. 5 for overview of Aim 2 results). For example, VSL-Direct was predicted by some aspects of both EDM and IPM. However, it is important to note that the relationship between VSL-Direct and multiple memory systems was inconsistent (e.g., VSL-Direct was related to Paired Associated, but not Visual Object Learning). In addition, instructional condition affected the relationship between VSL-Indirect and executive function performance. Furthermore, in the explicit condition, executive function (attention) performance positively predicted VSL-Indirect performance. Yang and Li (2012) found differences in the recruitment of brain regions related to EDM, IPM, and executive function due to instructional condition. The authors found differential patterns of cortical-subcortical

connections. In the implicit condition, there was a direct cortical-subcortical connection (IFG to Caudate) during learning. However, in the explicit condition, this connection was mediated by the insula, a region related to attention/attentional networks (see Menon et al., 2010 for a review). Their behavioral data supported this finding, such that in the explicit condition, working memory (related to attention/executive function) predicted SL performance, but this relationship disappeared in the implicit condition.

However, unlike the behavioral findings from Yang and Li (2012), in the current study there was a trade-off between executive function skill and VSL-Indirect in the implicit condition. These findings are consistent with Shekiela et al. (2016) who found that concurrent working memory/executive function tasks interfere with SL performance (with implicit instructions) across visual and audio modalities. This suggests increases in executive function activity makes SL more difficult in implicit instructional conditions (i.e., a trade-off). As executive function and EDM are strongly related and have some shared circuitry, these findings are also consistent with Poldrack (2001). It is possible that increased executive function/attention activity in individuals with higher levels of executive function performance may disrupt the direct cortical-subcortical connection present during SL with implicit conditions (Yang and Li, 2012). Furthermore, the interference of executive function performance on SL in the implicit condition may be indicative of the competitive nature of EDM and IPM networks during the learning processing (Poldrack et al., 2001).

In conclusion, while there were no differences in performance across groups, the current results provide important insights into the componential nature of SL (domain general calculations) and the relationship between SL and multiple memory systems. First, performance on the VSL is related to EDM, IPM, EF, and vocabulary regardless of

learning condition, contrary to the assumption that SL is strictly an IPM process (see Saffran et al., 1997; Aslin and Newport, 2004; Conway and Christiansen, 2006; Perruchet and Pacton, 2006). In addition, direct and indirect (explicit, implicit) measures of VSL performance reacted differently to the instructional condition. The processes underlying the direct measure seemed to be unaffected by instructions. However, the instructional manipulations generated differences in the processes underlying indirect measure (correlation with EDM, trade-off with EDM in explicit and implicit respectively).

However, as mentioned previously, there were several cases in which there were marginal interactions with instructional condition. Like with the relationship between VSL-Direct and VSL-Indirect, the numerical or marginally significant interactions hint at the potential for additional insights that are occluded by sub-optimal measures. For example, with VSL-Indirect, there was a marginally significant interaction between instructional condition and visual object learning (EDM) such that in the implicit condition there was a significant negative relationship, but the relationship disappeared in the explicit condition. This pattern would match the relationship between VSL-Indirect and EF and provide a stronger conceptual link between EDM and EF. On the other hand, some of the findings discussed previously may be spurious or simply the result of noise. More sensitive measures are needed to more fully explore the nature of the relationship between VSL and multiple memory systems.

Exploratory Aim 3. Exploring Potential Improvements to SL and IPM Measurement

Aims 1 and 2 used established measures (VSL, IPM, EDM, etc.) from the literature. Several strict criteria were used to select the most appropriate measures. In recent years, various legitimate methodological issues with VSL and IPM measures in particular have

been identified and attempts have been made to develop better tasks and methods for analyzing the data to help mitigate these concerns. As discussed previously, while there are methodological issues, these measures should not be discarded as they have a strong theoretical foundation and a large body of supporting neurocognitive evidence (brain and behavior). Instead, Aim 3 sought to leverage advances in analysis and processing techniques to develop new methods of VSL and IPM measurement using advanced statistical analysis which were potentially more methodologically and theoretically sound.

The VSL task in particular has documented issues with reliability (test-retest and split-half), particularly with children (Anbal, 2019). In addition, within each task (VSL-Direct, VSL-Indirect, AGL, SRT) there were many nuisance variables due to the structure of each task that may affect performance and/or introduce additional noise, such as the presentation order of the target-foil pairs in the VSL or RT to previous items as we might see in a typical lexical decision task. This is particularly important as these measures were originally developed for group-level studies to establish the theoretical link to each specific memory/learning system (VSL, IPM). They were then later adapted for individual differences analyses. The shift from group-level to individual differences makes each of these measures more susceptible to nuisance variables that may inject noise at an individual level. Using advanced analyses provided an opportunity to advance these measures to provide a more accurate picture of individual differences in each construct. For example, controlling for these types of variables may help provide a more precise measure of individual differences. Lastly, some of the current versions of the measures may occlude or omit potentially interesting/important information gathered such as

amount of exposure during the SRT (learning curve, development of predictability effect over time).

Aim 3 sought to develop improvements to VSL (VSL-Direct, VSL-Indirect, Serial Reaction Time, Artificial Grammar Learning) measurement using advanced statistical analyses. In order to develop improved measures, several considerations needed to be addressed. First, the measures were assessed to examine whether improvements were needed. There are relatively few in-depth analyses of the psychometric properties of the chosen VSL and IPM measures in the literature. The current study provided an unprecedented opportunity to explore the distributional properties and reliability of several SL and IPM measures simultaneously with a relatively large sample size. This first consideration is the most important, as if the established mean measures had acceptable reliability, there would be no need to continue with applying advanced statistical methods and developing additional measures. There are a specific set of drawbacks related to the assumptions and calculations underlying each statistical method proposed in the following section (e.g., shrinkage/taking group data into account in LME potentially decreasing task validity). Therefore, the simplicity of the method (e.g., fewer assumptions, transformations) was also taken into account.

After the need for improvements was established, several major methodological and theoretical considerations were addressed. First, specific nuisance variables that may inject unwanted noise into the signal were accounted for. In addition, the established means/difference scores used in Aims 1 and 2 were examined to see whether they occlude interesting sources of information such as the effect of the amount of experience on the measure (i.e., learning effect). Relatedly, some tasks (e.g., AGL) had additional

measures which could be derived from the original data. These measures were then examined as described above (validity, reliability) and compared to the more established measures.

Once the measures were developed based on these criteria, they were evaluated. These measures were evaluated on a theoretical level (e.g., what added value does this measure provide?) and on a methodological level. Specifically, they were examined regarding their reliability (random-sample split-half) and measurement validity (e.g., theoretical considerations, convergent validity, divergent validity). In addition, the distributions of each of the measures were examined.

Aim 3 Method

For each established mean or mean difference measure, the psychometric properties of the established measures were examined (e.g., distributional information, reliability). For example, the split-half reliability of each measure was calculated for each participant (random sampling). This process was repeated at least 1000 times for each task to provide a mean and standard deviation for the reliability of each of the measures. Then to develop the improved set of measures for each task, mixed effects and regression models were used. Using the considerations in the previous step, Mixed Effects models (linear or logistic depending on the measure) were constructed.

The fixed effects structure of the LME models provided initial evidence that the considerations may improve the measures (e.g., a significant effect of trial order or previous trial type that needs to be statistically controlled). Furthermore, the random effects structure of the model confirmed whether the effect of the proposed variables differed at an individual level, therefore injecting noise into the individual differences

measure. Individual intercepts (for established mean measures) or coefficients (for established mean difference scores) were extracted from the completed LME model to serve as individual differences measures. In addition, as LME models have concerns with validity due to issues such as shrinkage, individual regression models based on the significant random effects from the LME model were generated and individual effect intercepts and coefficients were extracted for use as additional individual difference measures.

Lastly, the psychometric properties of each of the newly generated measures were evaluated. In particular, the reliability and validity (convergent, divergent) of each of these measures were examined. It is important to note that for each repetition of the random sample, all versions of the measures (e.g., basic mean, individual regression, LME) were calculated in order to enable more direct comparisons of reliability.

Aim 3 Results

Correct response times (RTs) and accuracy rates (ACCs) were analyzed using linear (or logistic) mixed effects (LME) modeling in R (Baayen et al., 2008). Subjects and items were entered as crossed-random factors. Reaction time (RT) data were log transformed prior to entry into the model to approximate a normal distribution. Analysis of reaction time generated t-values. An absolute t-value near two is considered an appropriate indicator of significance (see Baayen et al., 2008 for a review). Additionally, following the procedure outlined in Baayen et al. (2010), significance was determined by examining change in chi-squared. Examining delta chi-squared enables one to examine whether a variable explained a significant amount of variance in the model (improved the fit of the model). The analysis of error rates was conducted using the binomial function,

which generates z scores from which p values could be directly calculated. The LME coefficient, b, is reported for the effects of interest to provide insight into the relationship between the fixed effect factor and dependent variable, along with the standard error. The individual random effects structure was established using the log-likelihood ratio model comparison test and included participant and item as intercepts.

Table 5. Mean and Standard Deviation of VSL & IPM Measure Reliability (Pearson Correlation)

Construct	Measure	Subcomponent	Measurement/Analysis Type	Mean	SD	
VSL (Direct/Explicit), Statistical Control	VSL-Direct, All Data		Mean	0.73	0.02	
			Logit*	0.75	0.02	
			LME intercept	0.76	0.02	
	VSL-Direct, 2AFC Only		Mean	0.59	0.03	
			Logit	0.64	0.034	
			LME Intercept	0.61	0.03	
			Ind Reg Intercept	0.25	0.1	
VSL (Indirect/Implicit), Statistical Control	VSL-Indirect	Difference	0.1	0.05		
		LME Coefficient*	0.55	0.16		
		Ind Reg Coefficient	0.09	0.06		
Implicit/Procedural (IPM), Statistical Control	SRT, Blocks 3-8		Difference	0.22	0.06	
			Difference	0.26	0.06	
	SRT, ALL		LME Coefficient*	0.52	0.09	
			Ind Reg Coefficient	0.3	0.05	

Implicit/Procedural (IPM), Statistical Control	AGL, Established (Grammaticality Judgment)	Grammaticality	<i>dPrime*</i>	0.4	0.07
			Mean	0.28	0.05
			LME Intercept	0.28	0.05
Implicit/Procedural (IPM), Additional Measures	AGL, Endorsement	Grammaticality	<i>LME Coefficient*</i>	0.35	0.05
			Ind Reg Coefficient	0.19	0.11
		Clustering	<i>LME Coefficient*</i>	0.34	0.06
			Ind Reg Coefficient	0.17	0.1
	AGL, Endorsement, Ungram	Clustering	LME Coefficient	0.28	0.07
			Ind Reg Coefficient	0.19	0.09

VSL-Direct. Measurement Selection: Logit Transformation (VSL-Direct-Logit)

Psychometric properties of established measure (VSL-Direct). Table 5 shows the mean, standard deviation, and maximum and minimum reliability score for all of the variants of VSL and IPM measures. As described in the previous section, individual performance on VSL-Direct had a slightly right-skewed distribution, but no floor or ceiling effects. In addition, results of the split-half reliability found that the measure had moderate-high reliability ($M=.73$, $SD=.02$).

Methodological and Theoretical Considerations. While the established version of the measure has reasonable reliability, several factors may introduce noise in the established version and were statistically controlled for. For example, although VSL-Direct is considered a post-learning/exposure measure, participants were exposed to additional repetitions of complete, correct triplets throughout the test phase. In addition,

the presentation order of the target-foil pairing on 2AFC trials may affect performance on the current triplet and on the following triplet (e.g., if the participant sees the target second, their answer on the following item may be biased). Lastly, in the pattern completion portion of the task, the location of the missing item (second or third position) and of the target item (to complete the pattern) may affect performance due. In addition, the previous location of the missing item and target may affect performance in a similar way to the 2AFC trials.

Statistically Controlling for Nuisance Variables. In order to statistically control for these nuisance variables, both LME and individualized regressions were used. This approach was used for all subsequent tasks. First, an LME model was used to examine group-level effects and explore the random effects structure. Unfortunately, due to the nature of the various trial types in VSL-Direct, including factors which do not have values for some trials removes these trials. Overall, for all trials combined (2AFC, pattern completion), trial order was not significant. In addition, only the previous presentation order (2AFC trials only) was significant in the model, masking out the pattern completion trials.

Turning to individual differences, there were two possible directions for using LME to generate measures. First, to include all of the trials, a very limited LME model was used which only included the random effects of triplet and participant (note: this cannot be replicated in the regression model as the model would be underspecified). Secondly, the pattern completion trials were taken out and the 2AFC (64 trials) were focused on to statistically control for presentation order.

In the first case, individual intercepts for each participant were extracted. Reliability on this measure was slightly higher than for the basic measure ($M = .76$, $SD = .20$) and this difference was statistically significant ($t(1957) = 104.85$, $p < .001$). However, the reliability of the established mean measure simply transformed into logits (on the same scale as the intercepts extracted from the LME model) and the reliability of the LME coefficients was nearly identical ($M = .75$, $SD = .20$). Much of the increase in reliability may simply be due to the transformation to log odds. Logit transformations have been used on proportion correct outcome measures to facilitate individual differences research in language processing (e.g., Mirman, 2014). Probability is bounded 0 to 1 but the logit transformation is not. The transformation to logits therefore allows for more differentiation particularly at more extreme values (close to 0 or 1). Therefore, the transformation of proportion correct to logits without using the LME model may be a better option than extracting individual intercepts from the LME model in this particular case given there were no significant random effects.

The 2AFC-only results provided a slightly different picture. As a baseline, the reliability of the mean proportion correct was examined. Unsurprisingly, as 32 three alternative forced choice trials (pattern completion) were removed from the analysis, the mean scores from the 64 remaining 2AFC trials ($M = .59$, $SD = .03$) were less reliable than the mean from all 96 trials. This is consistent with Siegelman et al. (2014) in which the reliability of their VSL-Direct measure increased with the inclusion of trials of varying difficulty. The 2AFC-only measure had both a lower mean reliability score and also higher variability. Turning to the LME version of the model, previous presentation order was significant as a fixed effect ($b = .16$, $SE = 0.04$, $|z| = 3.85$, $p < .001$). However, entering

this variable as a random effect did not improve the fit of the model ($\Delta\chi^2 = 1.70$, $p > .05$). This suggested the LME model would not actually contribute to cleaning up noise at the individual level. In addition, as was the case with the initial version of the model, reliability on the LME version was statistically significantly higher than on the mean measure ($M=.61$, $SD = .003$) but was lower than a simple logit transform of the basic measure ($M=.64$, $SD=.034$).

Alternatively, as LME causes shrinkage to the mean of the distribution (more extreme values shift closer to the group mean) and may affect individual differences analyses, individualized regression models can be used to remove this effect. Individual regression models were based on the random effects structure of the LME analyses. The reliability of the individual intercepts generated by the individual regression models was significantly lower ($M= .25$, $SD= .01$) than both the basic mean score (2AFC only) and the LME version.

It is important to note that all of the additional measurements described had roughly normal distributions with no apparent floor or ceiling effects.

VSL Score Measurement Selection: Logit Transformation. In generating the best possible method for examining a construct, both the reliability and psychometric properties of the measure must be balanced with theoretical constraints and the overall simplicity of the method. For example, the basic mean score has reasonable reliability and a useable distribution. However, several theoretical and methodological factors may be considered in selecting the best possible method. While both the established mean score and the limited LME model have the highest reliability, the simple logit transformation of the established mean score had a similarly high reliability without the

issues stemming from shrinkage in LME. Therefore, measures that balances reliability, validity, and simplicity best is the logit transformation.

VSL-Indirect. Measurement Selection: LME Coefficients (VSL-Indirect-LME)

Psychometric properties of basic measure. Table 5 shows the mean and standard deviation of the reliability scores for all of the VSL-Indirect measures. Similar to the VSL-Direct measure, VSL-Indirect has a normal distribution with no floor or ceiling effects. However, because it is a difference score, the measure is inherently noisy. This is reflected in the very low reliability of the basic difference measure ($M = .1$, $SD = .05$).

Methodological and Theoretical Considerations. Several factors may be affecting the reliability of this measure and introducing noise. First, as this measure was placed directly after Exposure and items were presented one at a time in triplet order (as in the Exposure phase), VSL-Indirect may be acting as a secondary Exposure phase. Participants may be learning throughout the course of the experiment and amount of exposure should be accounted for as individuals may differ in the rate they learn. In addition, like with 2AFC trials in VSL, presentation order may bias performance. Four full triplets were presented in each trial. The target item may be present in any of the last three triplets. The triplet order may affect performance as items later in a trial may be responded to slower due to the need for sustained attention throughout or responded to faster as participants have additional information as to the location of the item as the trial continues. Additionally, previous presentation order may bias participants to expect a

target at a specific time in the sequence. Similarly, previous trial reaction time may affect current trial reaction time.

Statistically Controlling for Nuisance Variables. Each of the factors discussed above were significant within the model suggesting each of these variables had an effect on RT at the group level. Reaction times on items later in the trial (presentation position 3, 4) were responded to faster than items earlier in the trial (presentation order 2). This may be due to the fact that later in the trial, participants have more information and implicitly know that there are fewer options. Furthermore, this finding was reversed for the effect of the previous trial. The earlier the target was in the previous sequence, the faster the RT on the following trial. Interestingly, throughout the task, participant RTs actually increased. Once the nuisance variables were statistically controlled for, the effect of item in triplet could be examined (equivalent of looking at the RT difference between predictable and unpredictable items). The effect of predictability was statistically significant ($b = -0.068$, $SE = 0.005$, $|t| = -12.56$, $\Delta\chi^2 = 219.91$, $p < .001$) such that predictable items (item 3 in a triplet) were responded to faster than unpredictable items (Fig. 6).

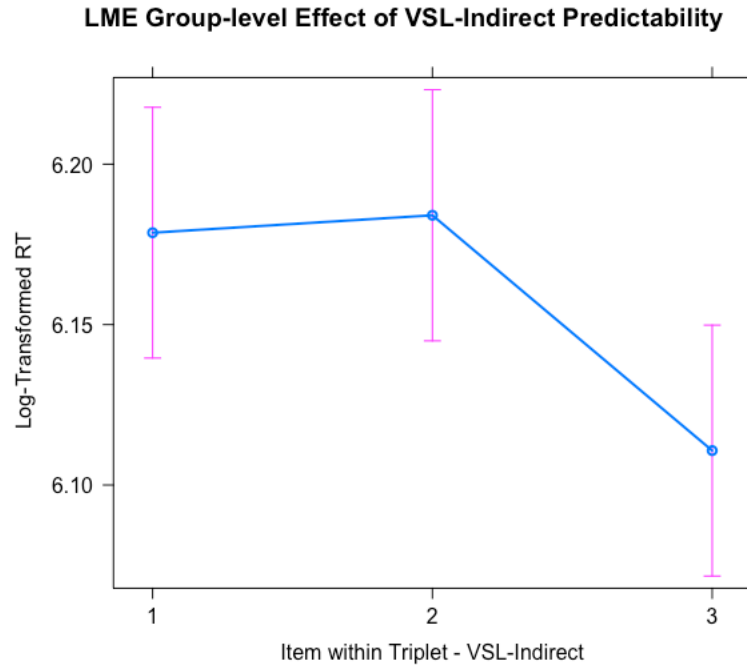


Figure 6. Plot of the effect of item predictability in the VSL-Indirect LME model.

Turning to individual differences, the random effect of predictability on participant was included in the LME model. Inclusion of this random effect significantly improved the fit of the model ($\Delta\chi^2 = 16.87$, $p < .01$), suggesting there were significant individual differences in the effect of predictability. In addition, inclusion of the random effects of presentation order ($\Delta\chi^2 = 31.83$, $p < .001$) and overall trial order ($\Delta\chi^2 = 31.40$, $p < .001$) also improved the fit of the model. This suggests that these variables may have introduced noise to the measurement and inclusion in the model will control for the individual effect of these variables.

Individual coefficients from each participant were extracted to examine individual differences. After controlling for the nuisance variables described above, the reliability of this measure increased significantly ($M = .55$, $SD = .16$). However, the individual regression model coefficients were interestingly more in line with the basic mean score

($M = .10$, $SD = .06$). Both the LME and individual regression coefficients were normally distributed.

Additional Measurements: Learning Rate. As stated above, individuals may differ to the degree that additional exposure in the VSL-Indirect affects performance. Within the group-level LME model, adding an interaction between trial and item accords for a significant amount of variability ($b = -.02$, $SE = 0.005$, $|t| = -2.40$, $\Delta x^2 = 13.99$, $p < .05$). At a group-level, overall RT increases throughout the experiment. However, the learning effect in VSL-Indirect develops as RT to unpredictable items (Item 1 in the triplet) slows rapidly throughout the experiment. Conversely, RT to predictable items (Item 3 in the triplet) is less affected by the change in overall RT and is able to maintain relatively fast RTs (Fig. 7).

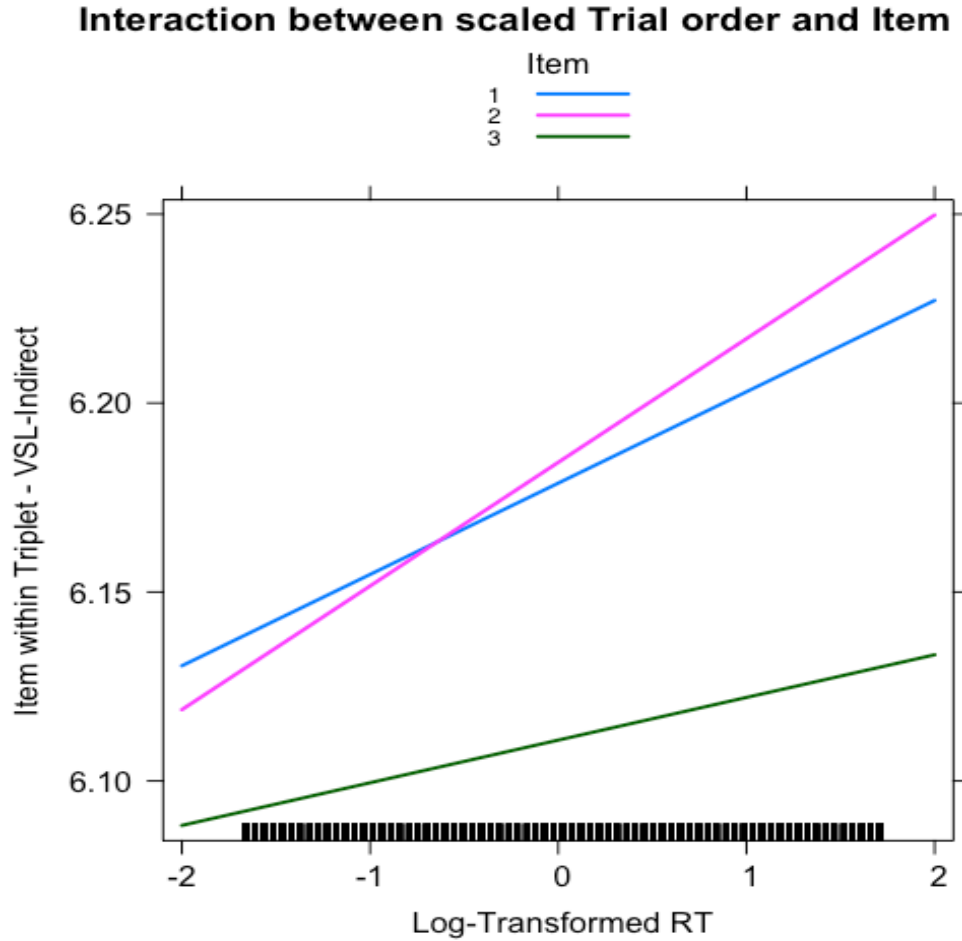


Figure 7. Plot of the interaction between Trial order (scaled) and Item in Triplet from the group-level LME model.

However, including the random effect of the interaction between trial and item did not improve the fit of the model ($\Delta\chi^2 = 16.59$, $p > .05$). It is important to note that attempting to include nonsignificant interactions (or any nonsignificant random effect from the LME) in an individual regression model to obtain an individual difference measure both generates very unreliable and theoretically unsound measures ($M = .004$, $SD = .05$) and lowers the reliability of the main measure (effect of predictability) ($M = .07$, $SD = .07$).

VSL-Indirect Measurement Selection: LME Coefficients. In this case, the best method for measuring VSL-Indirect is clear. The individual LME coefficients both have

the highest reliability and control for the theoretical and methodological concerns outlined above. Many factors related to presentation order and trial inject noise into the measure. In addition, as a difference measure, the basic difference score has an inherent issue with noise whereas the effect coefficients are slopes. Interestingly, including individual differences in learning rate (interaction between trial and item) did not improve the fit of the model even though there was a group-level effect present.

VSL Recommendations

Regarding the components of the VSL task, the current study found that: (i) the best method for analysis on the VSL-Direct measure was the simple Logit transformation; (ii) the best method for the VSL-Indirect measure was the individual LME coefficients; (iii) VSL-Indirect was more affected by nuisance variables (trial order, presentation position) than VSL-Direct and these factors need to be accounted for; (iv) overall the individualized regression models have much lower split-half reliability than the LME coefficients and the mean proportion correct/difference measures; and (v) the LME coefficients seemed to be more effective for increasing the reliability of the difference measure rather than the proportion correct.

Serial Reaction Time Task. Measurement selection: LME Coefficients (SRT-LME)

Psychometric properties of basic measure. Table 5 shows the mean and standard deviation of the reliability scores for all of the SRT variants. Like VSL-Indirect, the SRT difference score had a roughly normal distribution with no floor/ceiling effects, but had low reliability ($M=.22$, $SD=.06$).

Methodological and Theoretical Considerations. Unlike the two VSL measures described above, the serial reaction time task was not measured post-learning. The outcome measure is instead indexed during the learning phase and may provide additional information regarding the rate of learning. However, Kaufman et al. (2010) averaged RT across blocks 3-8 for predictable and unpredictable items in order to account for differences in learning rate overall. In addition, blocks 1 and 2 were not included in their original analyses because learning was not clearly established at the group level until block 3 so Kaufman et al. (2010) averaged across the final blocks of the experiment to get a more reliable signal.

In averaging across the final blocks, individual differences in learning rate was lost. For example, with this measure, one cannot examine the rate of learning. Additionally, while the learning effect is an average of the final blocks of the experiment (as at a group level there was not a clear learning signal in the initial blocks) there may be individual differences hidden within these initial blocks. Omitting the first two blocks may also miss the most important blocks for modelling individual differences in the speed of learning. In addition, learning at an individual level may differ and/or may be most reliable in the middle of the experiment or at the end.

Furthermore, there were several nuisance-type variables that may be controlled for to provide a cleaner signal. As Siegelman et al. (year) noted, the RT measure may have some issues with reliability. Controlling for some of these variables may help with these concerns of reliability. For example, as the experiment was set up in a block format there were several order effects one might consider controlling for, such as overall trial order (241-960), block order (1-8), and trials within each block (1-120). In addition,

participants may also have differences in reaction time due to the effect of items immediately preceding the current item. For example, participants may have slower RTs to items immediately following an improbable item or an item that they responded to very slowly or quickly. Controlling for the type of stimulus (e.g., left/right hand press, button press (1-4)) may also cut down on noise.

Statistically Controlling for Nuisance Variables. An initial LME model was used to explore the fixed and random effects structure and examine the effect of the various variables discussed above on RT. There was a significant effect of sequence structure (probability) in the model ($b = -.01$, $SE = 0.005$, $|t| = 4.67$, $\Delta x^2 = 12.64$, $p < .001$), confirming the RT difference results. In addition, block order, trial order, block x trial, and previous reaction time and trial type were significant in the model. Overall, throughout the experiment, RTs decreased. Furthermore, as previous RT increased, current RT also increased. Lastly, RTs after responding to items 1 or 2 were slower than RTs after responding to items 3 or 4. This is likely due to the fact that items 1-2 were responded to with the left hand and items 3-4 were responded to with the right hand.

Turning to the random effects structure, inclusion of probability, trial order, and previous trial location (1-4) improved the fit of the model. Each of these variables had an effect of RT at the individual level and introduced noise to the measure. Individual coefficients from each participant were extracted. As was the case in the VSL-Indirect and VSL-Direct, after controlling for the nuisance variables, the reliability of the LME coefficients ($M = .52$, $SD = .1$) was significantly higher than the basic difference measure. In addition, the individual regression models based on the LME random effects structure yielded reliability slightly higher ($M = .29$, $SD = .1$) than the mean difference scores, but

again lower than the LME coefficients. All of the additional measurements described had roughly normal distributions with no apparent floor or ceiling effects.

Additional Measurements: Learning Rate. Individuals may differ in learning rate. At the group level, there is an increase in the probability effect across blocks (see Figure 16). Differences in the rate of this increase may provide an additional measure of sensitivity to structure (i.e., a rate of learning). Within the group-level LME model, adding an interaction between block and probability accounts for a significant amount of variability ($b = -.01$, $SE = 0.005$, $|t| = 4.67$, $\Delta\chi^2 = 12.64$, $p < .001$) (Fig. 8). At the group level, the predictability effect develops throughout the course of the blocks. RTs to both predictable and unpredictable items get faster overall throughout the experiment, but RTs to predictable items decrease at a faster rate than unpredictable items.

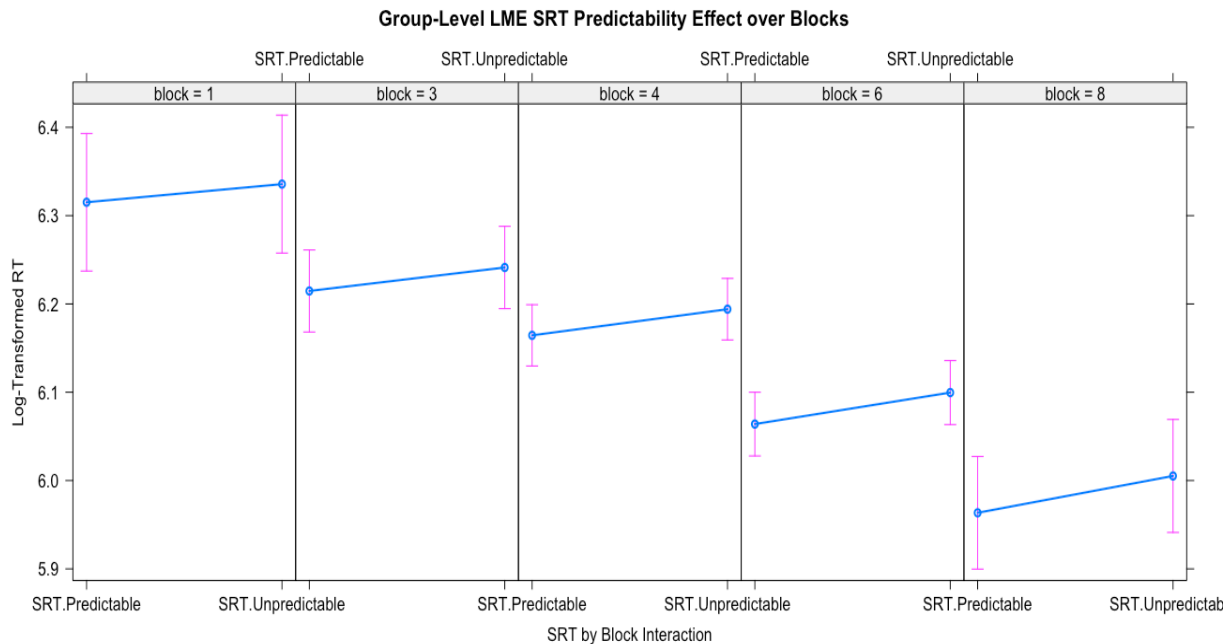


Figure 8. Plot of the interaction between block and SRT predictability from the group-level LME model.

However, like with the VSL-Indirect, including the random effect of this interaction did not improve the fit of the model. Furthermore, consistent with the VSL-Indirect, forced

inclusion of non-significant interactions in the individual regression models generates very unreliable measures (e.g., probability * block, $M=.02$, $SD=.06$) and lowers the reliability of the main measure (probability, $M=.03$, $SD=.04$).

SRT Measurement Selection. As was the case with the other difference measure (VSL-Indirect), the best method for measuring SRT was clear. The LME individual coefficients have both the highest reliability and controlled for the theoretical and methodological concerns outlined. Individual differences in changes in RT due to trial order and previous trial type introduced systematic noise that needs to be controlled in order to make the SRT more reliable. Interestingly, inclusion of the interaction with block in order to generate a measure of learning did not improve the fit of the overall LME model and were not useable.

Artificial Grammar Learning. Measurement Selection: dPrime (AGL-d'), LME Endorsement Rate (AGL-Gram, AGL-Clust)

Psychometric properties of basic measure. Table 5 shows the mean and standard deviation for all of the AGL measures. Both the proportion correct and the dprime measures have normal distributions with no floor or ceiling effects. The proportion correct has low-to-moderate reliability ($M=.28$, $SD=.05$) and the dprime measure has a significantly higher moderate reliability ($M=.4$, $SD=.05$).

Methodological and Theoretical Considerations. The AGL proportion correct measure has several methodological and theoretical issues. First, there are several factors which may introduce noise. For example, “grammaticality” and “clustering” were fully crossed, so getting a clean measure of “grammaticality” may be difficult without controlling for the effect of “clustering” post-analyses. Like with VSL-Direct, during the test

phase, participants are exposed to additional repetitions of correct and incorrect trials. There may be individual differences in the effect of this additional exposure. Further, the grammaticality and clustering-level of the previous trial may have an individual effect on responses to the current trial.

In addition to these issues with nuisance variables, the current version of the AGL has issues with overall validity. First, the underlying assumption of the AGL is that through exposure, participants learn the rules that govern the artificial grammar and are therefore able to recognize which sequences follow this grammar. However, ungrammatical items in this task (Pavlidou, 2017) are generated by changing one item in the sequence. Furthermore, the “grammatical” items consist solely of images participants have seen previously. There were no items that are grammatical but unfamiliar. The lack of unfamiliar grammatical items makes the assumption that participants are learning the “rules” difficult to support. As the current measure stands, “grammaticality” is more likely simply familiarity learned throughout exposure.

Additionally, clustering (differences in the frequency of bigrams and trigrams within the sequences) was crossed with grammaticality (familiarity). Pavlidou et al. (2017) found an interaction at the group-level between grammaticality and clustering such that ungrammatical items were harder to reject if they had high clustering. This suggests participants were also sensitive to the statistical regularities present at smaller grain sizes (potentially a sublexical unit). Extracting clustering may provide an additional interesting measure of AGL related to statistical learning. For example, Siegelman et al. (2018) found individuals differ to the degree they rely on global structure vs. local co-occurrences in a

self-paced VSL task. The authors further speculate that these computations may be supported by different memory systems.

Statistically Controlling for Nuisance Variables. Interestingly, only clustering was significant at the group level ($b = -.22$, $SE = 0.04$, $z = -5.6$, $\Delta\chi^2 = 32.10$, $p < .001$) such that trials with high clustering had significantly lower accuracy (Fig. 9). Trial order, previous grammaticality, and previous clustering were not significant at the group level. However, none of these factors improved the overall fit of the model as random effects.

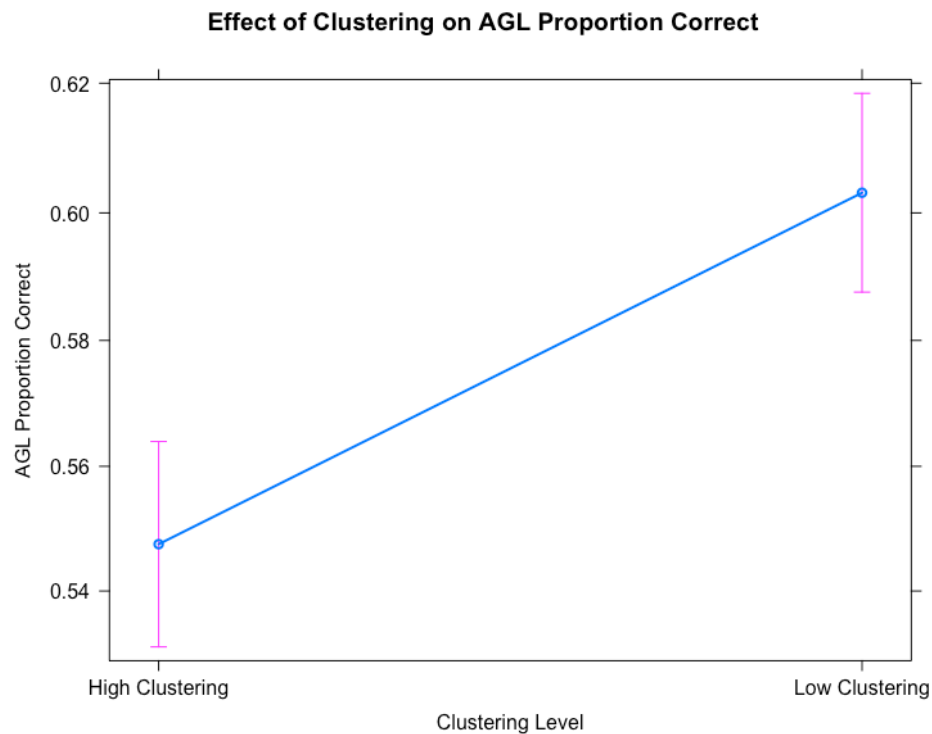


Figure 9. Plot of the effect of AGL Clustering on AGL Proportion Correct from the group-level LME model. High clustering items (both grammatical and ungrammatical) have high total bigram and trigram frequency (i.e., in exposure, items that have high levels of local co-occurrence).

Individual intercepts for the overall LME model were extracted. The reliability of this measure was nearly identical to the basic mean measure ($M=.28$, $SD =.05$). This is unsurprising as none of the nuisance variables entered as random effects. While there were significant group-level (fixed) effects, these effects shift all participants equally and

therefore do not correct any by-participant noise. As there were no significant random effects, individual regression models would be underspecified. However, if one were to regress the effect of clustering on accuracy in order to generate a model, both the intercept ($M = .18$, $SD = .09$) and the effect of clustering ($M = .07$, $SD = .06$) would have very low reliability as was the case with the inclusion of non-significant interactions in SRT or VSL-Indirect as random effects.

Additional Measurements: Endorsement Rate (Grammaticality, Clustering).

In order to examine sensitivity both grammaticality (familiarity) and clustering, endorsement rate may be used instead of proportion correct. Endorsement rate indexes the effect a specific variable has on the decision to select an item as correct (i.e., to endorse an item). Endorsement rate has been used in AGL for this exact purpose in many cases (e.g., Kinder, 2000; Folia et al. 2008). Endorsement rate in this case may be used to explore sensitivity to statistical structure at the more global-level (grammaticality, familiarity) and the local level (clustering) as Siegelman et al., (2019) did with VSL.

Endorsement rate may be affected by the same nuisance variables as the basic proportion correct measure, so they were included in the group-level model. Trial order and previous grammaticality were significant at the group level. Importantly, grammaticality ($b = .59$, $SE = 0.04$, $z = 14.65$, $p < .001$), clustering ($b = .29$, $SE = 0.04$, $z = 7.2$, $p < .001$), and the interaction between grammaticality and clustering ($b = -.48$, $SE = 0.08$, $z = -.594$, $p < .001$) were all significant at the group level (Fig. 10). Participants' endorsement rates decreased as the experiment went on. In addition, participants were less likely to endorse an item following a grammatical item. The interaction between clustering and grammaticality was consistent with the finding in Pavlidou et al. (2016)

such that there was a large effect of clustering for ungrammatical items but almost none for grammatical items (Fig. 11). This suggests, in order to obtain the cleanest measure of clustering, looking at only the ungrammatical items may be the best possible method.

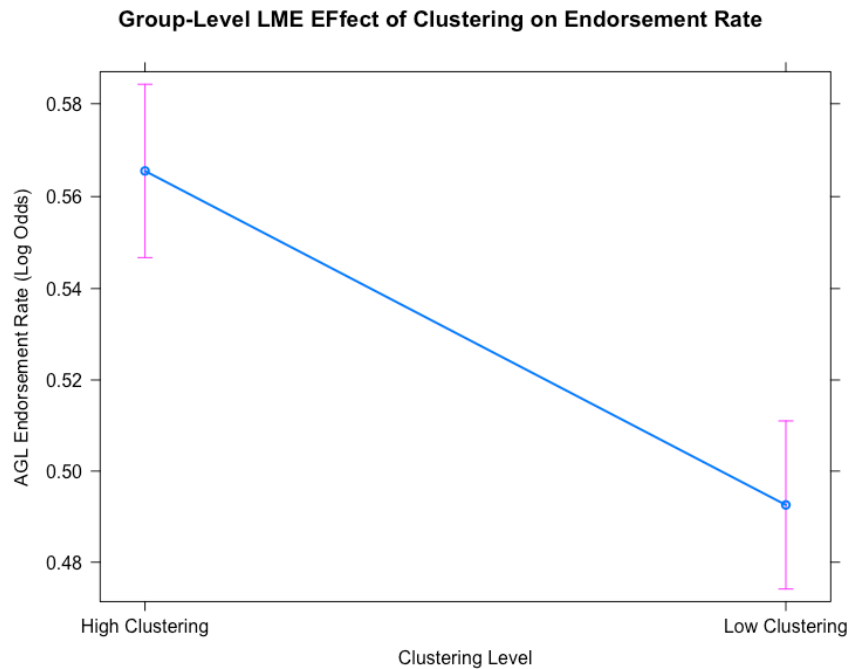
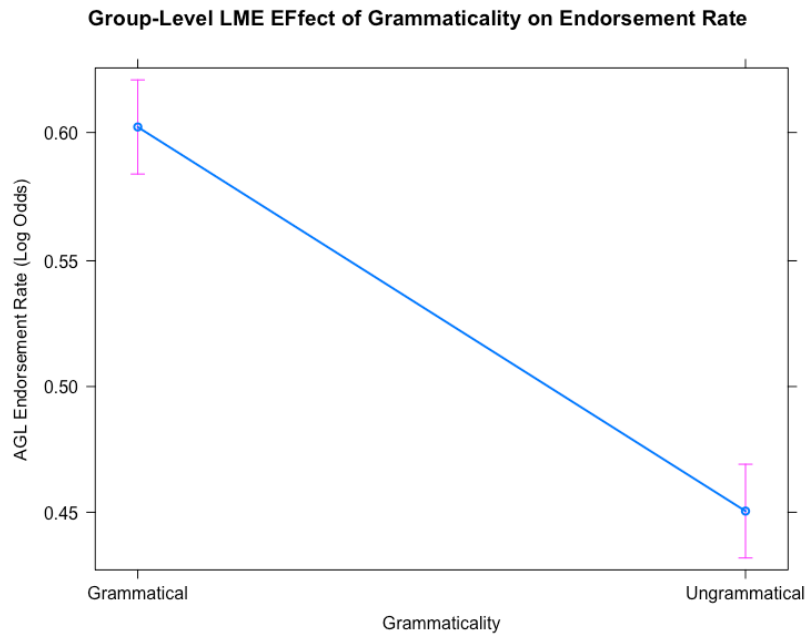


Figure 10. Plots of the effect of Grammaticality (top) and Clustering (bottom) on Endorsement Rate from the group-level LME model. High clustering items (both grammatical and ungrammatical) have high total bigram and trigram frequency (i.e., in exposure, items that have high levels of local co-occurrence).

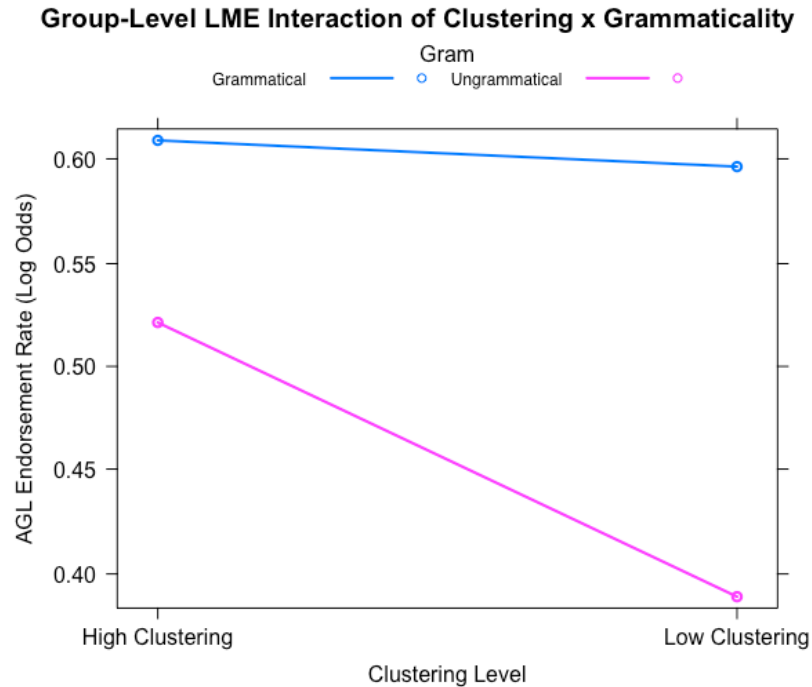


Figure 11. Plot of the interaction between Grammaticality and Clustering on Endorsement Rate from the group-level LME model. High clustering items (both grammatical and ungrammatical) have high total bigram and trigram frequency (i.e., in exposure, items that have high levels of local co-occurrence).

Turning to individual differences, only grammaticality ($\Delta x^2 = 36.35$, $p < .001$) and clustering ($\Delta x^2 = 40.87$, $p < .001$) improved the fit of the model as random effects. However, the interaction between grammaticality and clustering was not significant indicating individuals did not differ to the degree that the clustering effect was affected by grammaticality. As both variables were significant as random effects, they account for separate variance at an individual level and are separate predictors of endorsement rate (Xu and Taft, 2015). The individual coefficients for grammaticality and clustering were extracted from the model. After statistically controlling for the effect of grammaticality and clustering on each other, the reliability of each measure was significantly higher than the basic proportion correct, ($M = .35$, $SD = .05$) and ($M = .34$, $SD = .06$) respectively.

Consistent with the previous set of findings, the individual regression model coefficients were significantly less reliable than the LME coefficients, ($M = .20$, $SD = .11$) and ($M = .19$, $SD = .10$) respectively.

Additional Measurements: Endorsement Rate for Ungrammatical Items Only.

As noted previously, at the group level, there was a large effect of clustering for ungrammatical items, but an almost non-existent effect for grammatical items. In addition, this interaction was not significant as random effect, suggesting it was consistent across participants. Examining only unfamiliar/ungrammatical items may provide a cleaner measure of clustering as familiarity would not interact with sensitivity to local clusters and endorsement rate would be more driven by these local clusters and not more global memory traces.

After filtering out grammatical items, only 32 trials remain. However, the group-level effects (minus the interaction with grammaticality) were consistent. Furthermore, including clustering as a random effect improved the fit of the model ($\Delta\chi^2 = 25.16$, $p < .001$). Unsurprisingly, the extracted LME coefficients from the ungrammatical item only model were less reliable ($M = .28$, $SD = .05$) than with all 64 trials. The individual regression models were even less reliable ($M = .19$, $SD = .09$). While this measure is less reliable than the version using all trials, it may be a cleaner more valid measure overall. Ideally, there would be 64 ungrammatical trials in order to have a both reliable and valid measure. On the other hand, while there was little effect of clustering at the group-level, important individual differences may come out in these grammatical trials.

Note, all of the additional measurements described had roughly normal distributions with no apparent floor or ceiling effects.

AGL Measurement Selection: dPrime (AGL-d'), LME Endorsement Rate Grammaticality (AGL-Gram), LME Endorsement Rate Clustering (AGL-Clust). Unlike with the VSL and SRT, the best method for measuring AGL is not as clear. From a reliability perspective, the measure with the highest reliability is actually the dprime measure derived from the original data. However, as discussed previously, there are many issues inherent to measuring the effect of grammaticality (familiarity). If one is interested in examining individual sensitivity to grammaticality, the effect of grammaticality on endorsement rate (LME version) has a blend of acceptable reliability and a strong case for validity. Using this method also allows one to extract the effect of clustering, potentially a separate learning measure, with the effect of grammaticality controlled for. This method also accounts for all of the data, unlike with the ungrammatical-only case.

In conclusion, extracting the individual coefficients from the LME endorsement rate model for grammaticality and clustering provides the best blend of reliability and strong theoretical backing. Furthermore, in order to also include a grammaticality-judgement related, the dprime measure, which has the highest reliability score, may also be used. The dprime score, which comes from signal detection theory, is also a more valid measure in this case as participants make a accept/reject judgment on a single presented item (unlike the decision used in VSL). The dprime score is a score of discrimination between old/new (grammatical/ungrammatical) and controls for biases in responses.

Table 6. Measurement Recommendations

Measure	Subcomponent	Measurement/Analysis Type	Reasons for Selection
VSL-Direct, All Data	Direct	Logit transformation of the mean proportion correct	Relatively high reliability, fewer assumptions regarding post-analysis (shrinkage), simplicity of processing. Established proportion correct alternative in the literature.
VSL-Indirect	Indirect	LME Coefficient	Highest reliability of all variants. Uses slopes instead of a difference score. Controls for important nuisance variables at the individual level.
SRT, ALL	Predictability	LME Coefficient	Highest reliability of all variants. Uses slopes instead of a difference score. Controls for important nuisance variables at the individual level.
AGL, Established (Grammaticality Judgment)	Grammaticality	dPrime	Relatively high reliability, fewer assumptions regarding post-analysis (shrinkage), simplicity of processing. Established proportion correct alternative in the literature.
AGL, Endorsement	Grammaticality	LME Coefficient	Highest reliability of all variants. Controls for the impact of clustering at an individual level (cleaner measure of grammaticality). Endorsement rate allows for the analysis of sensitivity to regularities at multiple levels (local vs. global)

Clustering

LME Coefficient

Highest reliability of all variants. Controls for the impact of grammaticality at an individual level (cleaner measure of clustering). Endorsement rate allows for the analysis of sensitivity to regularities at multiple levels (local vs. global)

Exploratory Aim 3 Discussion

In Exploratory Aim 3, I examined the psychometric properties of VSL and IPM measures and attempted to develop better methods of data processing in order to address methodological and theoretical concerns identified in the literature. Exploratory Aim 3 provided several interesting insights regarding best practice for VSL and IPM measurement. Several interesting findings will be highlighted below.

1. LME coefficients had the highest reliability and Individualized coefficients had the lowest reliability.

Overall, an interesting pattern emerged regarding the reliability of the mean, individual regression, and LME measures. Within all four measures discussed (VSL-Direct, VSL-Indirect, AGL, SRT), the LME intercepts and coefficients had the highest reliability. While the individual regression models avoid validity issues regarding shrinkage, the reliability of these measures were many times so low they were unusable. This was particularly the case when interactions or variables that did not improve the fit of the LME model at a random effects level were included in the structure of the individual regression models.

One may argue that the distributions of the coefficients/intercepts from the LME and individual regression models generated from the random sampling may be highly skewed or have issues with outliers due to the split. However, when the reliability analyses are run using the Spearman correlation, the reliability results are nearly identical. This method is more robust to outliers and skewed data than Pearson correlations as the data are transformed into rank-order. In addition, the Spearman correlation also accounts for some potential non-monotonic relationships with the rank-order transformation. Lastly, Spearman correlation does not have the same assumptions regarding the meaningfulness of the distance between data points.

2. The low split-half reliability for the individual regression models may be explained by the fact that the number of trials were cut in half.

The fact that the individual regression model coefficients/intercepts are not more reliable than the LME coefficients/intercepts is unsurprising as the LME model accounts for both group and individual data (more data). However, both the individual regression models and the mean scores reference the same amount of data. The low split-half reliability for the individual regression models may be explained by the fact that the number of trials were simply cut in half. Due to this, the signal-to-noise ratio for each individual model may be decreased significantly (i.e., picking up only noise and no signal), severely reducing the reliability for the individual models.

3. LME coefficients provided a greater increase reliability over difference scores than LME intercepts provided over proportion correct.

The pattern of the LME coefficients having the highest reliability was magnified in the difference between the coefficients and the mean difference measures (SRT, VSL-Indirect). This is likely due to the fact that the LME coefficients corrects for both the nuisance variables and the noise inherent to difference measures. However, the benefit of using LME was less obvious for the basic means vs. LME intercepts (VSL-Direct, AGL Basic measure). This was particularly the case when the LME model did not have any significant random effects (no benefit of controlling for individual noise).

4. In some cases (particularly with the proportion correct measures) the simpler analysis approach produced the best result and should be balanced with reliability and validity.

In VSL-Direct, a simple transformation from proportion correct into log-odds (e.g., Mirman, 2014) improved the reliability of the measure as much as the LME model. This may be the preferred method, when the LME does not include any additional random effects. The simple logit transformation does not have the added issue of shrinkage (shifting of more extreme scores towards the group mean) or other artifacts related to accounting for group-level data. In addition, the increase in reliability is maintained. It is important to note that the simplicity of the method should be balanced with reliability and validity of the method in most cases (e.g., fewer assumptions, closer to the original data).

Therefore, both the VSL-Direct and AGL Grammaticality did not use LME intercepts, instead opting for more simple logit and dprime transformations.

Exploratory Aim 4. Re-examining the Findings from Aims 1 and 2

Exploratory Aim 4 sought to re-analyze Aims 1 and 2 using the updated methods developed in Aim 3 (see Table 6 for recommendations). The improved measures provided additional insights that were occluded using the established measures. These findings helped further develop the nuanced understanding of the relationship between VSL, EDM and IPM. Several questions were addressed:

- (i) **Was performance on VSL-Direct-Logit and VSL-Indirect-LME still identical across instructional conditions?**
- (ii) **Was the marginally significant interaction between VSL-Indirect and instructional condition (regressed onto VSL-Direct) confirmed or disconfirmed?**
- (iii) **Were paired associate learning (EDM), SRT-LME (IPM), Identification (executive function), and vocabulary still predictive of VSL-Direct-Logit?**
- (iv) **Were there any new predictors of VSL-Direct-Logit?**
- (v) **Was the marginally significant interaction between visual object learning-immediate and instructional condition (regressed onto VSL-Indirect-LME) confirmed or disconfirmed?**
- (vi) **Were there any new predictors of VSL-Indirect-LME?**
- (vii) **What were the characteristics of the newly developed measures (e.g., AGL-clust, AGL-gram)?**

(viii) How did these newly developed measures predict VSL performance?

Overall, the changes to the measurements made in Exploratory Aim 3 may help uncover additional insights into the relationship between VSL and multiple memory systems.

Method**Updated VSL and IPM Measures**

See Aim 3 (Table 6) for full descriptions of the VSL (VSL-Direct-Logit, VSL-Indirect-LME) and IPM (AGL-d', AGL-gram, AGL-clust, SRT-LME) used in Aim 4.

Other Measures

All of the EDM (Continuous Paired Associate Learning, Visual Object Learning, Visual Object Learning Delayed) and Cognitive Measures (Letter N-Back, Detection, Identification, Vocabulary) were the same as in Aim 2 (see Aim 2 methods for full task descriptions).

Results

As Exploratory Aim 4 was simply a re-examination of the findings in Aims 1 and 2, the results section focused on findings that: (i) were confirmed or disconfirmed, (ii) were new and expanded the understanding of the relationship between VSL and IPM and EDM.

Aim 1 Revisited

Confirmed: Performance on VSL was the same across instructional conditions

In VSL-Direct-Logit, performance again was nearly identical across groups. Further, in VSL-Indirect-LME instructional condition was not significant as a fixed effect and importantly, the interaction with item predictability was not significant. Entering these variables as random effects also did not improve the fit of the model. These findings suggested that instructional condition did not have an effect on overall performance.

Disconfirmed: The interaction between VSL-Indirect-LME and instructional condition (regressed onto VSL-Direct-Logit) was not significant.

To assess the relationship between VSL measures, a series of regression models were created. Like in Aim 1, VSL-Indirect-LME was not a significant predictor of VSL-Direct-Logit ($b = -1.94$, $t(204) = -.82$, $p > .05$). In addition, including the interaction between VSL-Indirect-LME and instructional condition did not improve the fit of the model ($F(2, 202) = 0.17$, $p > .05$).

Aim 2 Revisited**Table 7. Correlation Matrix of Updated Measures**

		Implicit/Procedural (IPM)				Explicit Declarative (EDM)			Cognitive Measures		
		AGL-d	AGL-g	AGL-c	SRT-L	VOL-I	VOL-D	CPAL	ID	DET	L-NB
IPM	AGL dprime	1.00									
	AGL-gram	0.73*									
	AGL-clust	-0.43*	-0.59*								
	SRT LME	0.10	0.05	0.10							
EDM	Visual Object Learning-Immediate	-0.04	-0.06	0.15*	-0.04						

	Visual Object Learning-Delayed	-0.01	0.01	0.03	0.02	0.57					
	Continuous Paired Associates	0.05	-0.06	0.17*	0.11	0.13	0.06				
	Identification	-0.09	-0.04	-0.05	-0.27*	0.04	-0.04	-0.10			
Cognitive Measures	Detection	-0.05	0.01	-0.02	-0.21*	-0.04	-0.04	-0.05	0.58*		
	Letter N-Back	0.00	-0.03	0.04	-0.15	0.01	0.00	-0.05	0.40*	0.35*	
	Vocabulary	-0.01	-0.05	0.24*	-0.11	0.28*	0.25*	-0.03	-0.11	-0.11	-0.09

Note: AGL dprime (AGL-d), AGL Endorsement: Grammaticality (AGL-g), AGL Endorsement: Clustering, SRT-LME (SRT-L), Visual Object Learning-Immediate (VOL-I), Visual Object Learning-Delayed (VOL-D), Identification (ID), Detection (DET), Letter N-back (L-NB)

Relationship between SRT and AGL

Relationship between IPM, EDM, and Cognitive Measures

Turning to the relationship between IPM and EDM, SRT-LME, AGL-d', and AGL-gram were not correlated with any of the EDM measures. However, AGL-clust was significantly correlated with both paired associate learning ($r = .17$, $p < .05$) and visual object learning-immediate ($r = .16$, $p < .05$). The correlation with EDM measures and the negative correlation with AGL-d' and AGL-gram may suggest that AGL-clust indexes separate aspects of information learned during the AGL task linked to EDM. Similarly, only AGL-clust was correlated with vocabulary. AGL-clust patterns with vocabulary in a similar manner as EDM. This would mirror findings in which individual performance increases with explicit instruction, but only for sequences in which the pattern is simple (rather than complex) (Frensch and Miner, 1994; Jimenez et al. 1996). Therefore, the relatively shorter, less complex regularities inherent in the bigrams and trigrams (relative to the more complex co-occurrences present in grammaticality) may be differentially

supported by EDM. It is important to note that the strength of correlation between AGL-clust and the EDM measures was very low and, given the number of comparisons, may in fact simply be spurious.

Lastly, consistent with Aim 2, only SRT-LME was correlated with any of the executive function measures. In addition to the significant correlation with detection found in Aim 2 ($r = -.21$, $p < .05$), SRT-LME was also correlated with identification ($r = -.27$, $p < .01$), such that individuals with faster reaction times for both IDN and DET (i.e., better EF) had larger SRT-LME predictability effects.

Confirmed: VSL-Direct-Logit was predicted by paired associate learning (EDM), SRT-LME (IPM), Identification (executive function), and vocabulary.

To examine the relationship between VSL and EDM across instructional conditions, regression analyses were conducted. The same method of model comparison from Aims 1 and 2 were used. All of the following regression analyses used this structure. First, to follow up the results from Aim 2, in Model 1.1, paired associate learning was regressed onto VSL-Direct-Logit. The results confirmed the findings in Aim 1. As in Aim 1, paired associate learning significantly predicted VSL-Direct-Logit ($b = .14$, $t(201) = 2.07$, $p < .001$). In Model 4.2, the interaction between paired associate learning and the instructional manipulation was added to the model as a predictor. The interaction term did not improve the fit of the model ($F(2, 199) = .33$, $p > .05$) suggesting the instructional condition did not affect the relationship between VSL-Direct-Logit and paired associate learning.

To examine the relationship between VSL and IPM across instructional conditions, regression analyses were conducted. First, to follow up the results from Aim 1, in Model

2.1, SRT-LME was regressed onto VSL-Direct-Logit. As in Aim 1, SRT-LME significantly predicted VSL-Direct-Logit ($b = 9.65$, $t(146) = 2.07$, $p < .05$). The interaction between SRT-LME and instructions (Model 2.2) did not significantly improve the fit of the model ($F(2, 144) = .50$, $p > .05$).

In Model 3.1, Identification was regressed onto VSL-Direct-Logit. Identification significantly predicted VSL-Direct-Logit ($b = -1.58$, $t(204) = -2.35$, $p < .05$). As was the case in Aim 2, the interaction between identification and instructional condition (Model 3.2) did not improve the fit of the model ($F(2, 203) = .56$, $p > .05$) suggesting the relationship between identification and VSL-Direct-Logit was not affected by the instructional condition.

Furthermore, in Model 4.1, Vocabulary was regressed onto VSL-Direct-Logit. As in Aim 2, Vocabulary was a significant predictor of VSL-Direct-Logit ($b = .18$, $t(142) = 2.45$, $p < .05$). The interaction between Vocabulary and instructions (Model 4.2) did not improve the fit of the model ($F(2, 140) = .29$, $p > .05$) suggesting that the relationship between VSL-Direct-Logit and Vocabulary did not shift due to instructions.

Confirmed: VSL-Direct-Logit predicted Vocabulary controlling for EDM

To confirm the finding in Aim 2 in which VSL-Direct was a significant predictor of vocabulary, controlling for EDM, multiple regression was used to clarify the relationship between these measures. First, in Model 5.1, VSL-Direct-Logit was regressed onto Vocabulary. VSL-Direct-Logit was a significant predictor of Vocabulary ($b = .04$, $t(142) = 2.45$, $p < .05$). In Model 5.2, visual object learning-immediate was regressed onto Vocabulary and was a significant predictor ($b = .46$, $t(153) = 3.58$, $p < .001$). In Model 5.3, visual object learning-delayed was regressed onto Vocabulary and was also a significant

predictor ($b = .37$, $t(151) = 3.14$, $p < .001$) as would be expected. In Model 5.4, both visual object learning immediate and delayed were regressed onto Vocabulary. Interestingly, only visual object learning-immediate was a significant predictor ($b = .35$, $t(148) = 2.29$, $p < .05$). Lastly, in Model 5.5, VSL-Direct-Logit and both visual object learning measures were regressed onto Vocabulary. After controlling for visual object learning (immediate, delayed), VSL-Direct-Logit was still a significant predictor in the model ($b = .04$, $t(132) = 2.06$, $p < .05$).

New Finding: AGL-clust was a significant predictor of VSL-Direct-Logit

Turning to the relationship between AGL-clust and VSL-Direct-Logit, in Model 6.1, AGL-clust was regressed onto VSL-Direct-Logit. AGL-clust significantly predicted VSL-Direct-Logit ($b = .43$, $t(145) = 2.77$, $p < .01$). The interaction between AGL-clust and instructional condition (Model 6.2) did not significantly improve the fit of the model ($F(2, 143) = .24$, $p > .05$). These results suggested that AGL-clust was related to VSL-Direct-Logit, but the instructional condition did not significantly shift processing.

Confirmed: VSL-Indirect-LME was predicted by visual object learning-immediate (EDM) and detection (EF) differently across instructional condition

In Model 7.1, visual object Visual Object Learning-immediate was regressed onto VSL-Indirect-LME. Like in Aim 2, Visual Object Learning-immediate was not a significant predictor of VSL-Indirect-LME ($b = -.02$, $t(149) = -1.36$, $p > .05$). In Model 7.3, the interaction between visual object Visual Object Learning-immediate and instructional condition was added to the model. Unlike in Aim 2, the interaction term significantly improved the fit of the model ($F(2, 147) = 4.42$, $p < .05$). To confirm this effect, an intermediate model (Model 7.2) was created which added the instructional manipulation

as a predictor but not the interaction with visual object learning-immediate. Then Models 7.1, 7.2, and 7.3 were compared. Instructional condition (Model 7.2) did not improve the fit of Model 7.1 ($F(1, 147) = 2.36$, $p > .05$). However, the including the interaction term improved the fit of the model over Model 7.2 ($F(1, 147) = 6.49$, $p < .05$). In the explicit condition, as visual object Visual Object Learning-immediate performance increased, VSL-Indirect-LME increased. However, in the implicit condition, as Visual object learning-immediate performance increased, VSL-Indirect-LME actually decreased.

Turning to the relationship between VSL-Indirect-LME and executive function, the results from Aim 4 confirmed the relationship between VSL-Indirect and detection found in Aim 2. In Model 8.1, Detection was regressed onto VSL-Indirect-LME. As was the case in Aim 2, Detection was a significant predictor of VSL-Indirect-LME ($b = .05$, $t(203) = 3.85$, $p < .001$). The interaction between Detection and instructional condition (Model 8.3) significantly improved the fit of the model ($F(2, 201) = 4.41$, $p < .05$). In the intermediate Model 8.2, the interaction term was removed (only including instructional condition as described previously). Instructional condition (Model 8.2) did not improve the fit of the model over Model 8.1 ($F(1, 202) = 1.47$, $p > .05$), but the interaction between detection and instructional condition (Model 8.3) improved the fit of the model over Model 8.2 ($F(1, 201) = 7.36$, $p < .001$). In the implicit condition, as Detection RT decreased (i.e., better EF), VSL-Indirect also decreased. This suggests that in the implicit condition, EF performance is actually detrimental to learning in the VSL-Indirect. In the explicit condition, there was no relationship between VSL-Indirect and Detection.

New Finding: VSL-Indirect-LME was predicted by visual object learning-delayed

Expanding on findings from Aim 2, the results for Visual object learning-delayed were nearly identical to visual object Visual Object Learning-immediate. In Model 9.1, Visual object learning-delayed was regressed onto VSL-Indirect-LME. Visual object learning-delayed was also not a significant predictor of VSL-Indirect-LME ($b = -.02$, $t(146) = -1.43$, $p > .05$). However, the interaction between visual object Visual Object Learning-immediate and instructional condition (Model 9.3) significantly improved the fit of the model ($F(2, 144) = 3.65$, $p < .05$) (Fig. 12). In addition, as with visual object Visual Object Learning-immediate, the intermediate model (Model 9.2), did not improve the fit of the model ($F(1, 145) = 1.92$, $p > .05$), but the interaction between visual object Visual Object Learning-immediate and instructional condition (Model 9.3) improved the fit of the model over Model 9.2 ($F(1, 144) = 5.38$, $p < .05$). As with visual object Visual Object Learning-immediate, in the explicit condition, as visual object Visual Object Learning-delayed performance increased, VSL-Indirect-LME increased. However, in the implicit condition, as visual object Visual Object Learning-delayed performance increased, VSL-Indirect-LME actually decreased.

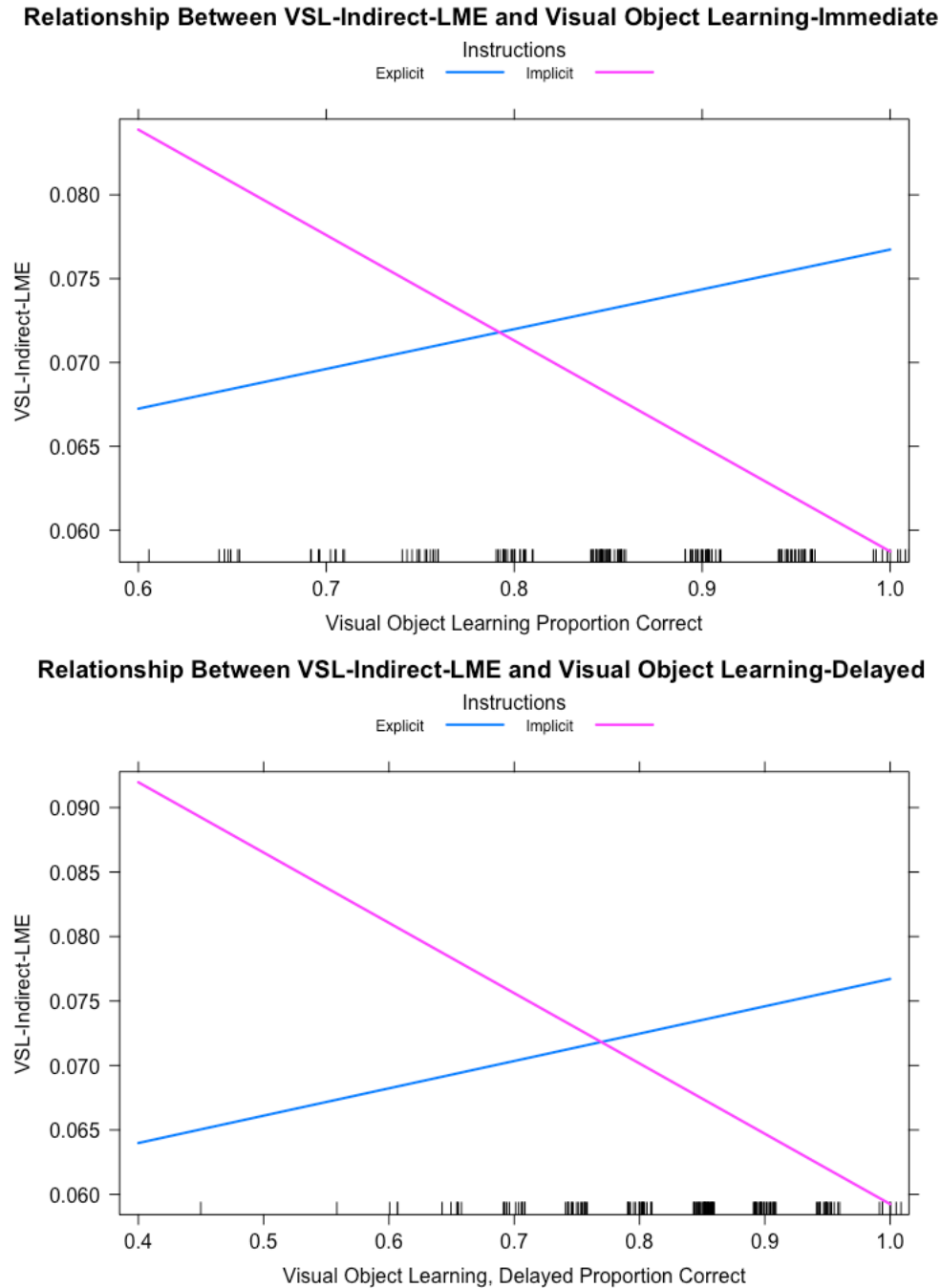


Figure 12. Relationship between VSL-Indirect LME and Visual Object Learning and Visual Object Learning – Delayed across the Instructional Manipulation.

New Finding: VSL-Indirect-LME was predicted by additional executive function measures (identification and letter n-back)

Similar to the results with detection (Aims 2 and 4), in Model 10.1, Identification was regressed onto VSL-Indirect-LME. Identification was a significant predictor of VSL-Indirect-LME ($b = .10$, $t(204) = 5.32$, $p < .001$). The interaction between Identification and instructions (Model 10.2) did not significantly improve the fit of the model ($F(2, 202) = 2.12$, $p > .05$). However, in Model 11.2, interaction between Identification and instructions was marginal ($b = .07$, $t(202) = 1.78$, $p = .076$) and the pattern mirrored the pattern present with Detection (Fig. 13). Overall, across all both executive function measures, as executive function performance increased, VSL-Indirect performance decreased. This effect was increased in the implicit condition as seen with the significant and marginally significant interactions with Detection and Identification respectively.

Turning to the relationship between VSL-Indirect-LME and working memory, in Model 11.1, letter N-back was regressed onto VSL-Indirect-LME. Letter N-back was a significant predictor of VSL-Indirect-LME ($b = .08$, $t(143) = 4.30$, $p < .001$). The interaction between Letter N-back and instructions (Model 11.2) did not significantly improve the fit of the model ($F(2, 141) = 2.25$, $p > .05$). In addition, in Model 12.2, the interaction between Letter N-back and instructions ($b = .001$, $t(141) = 1.61$, $p > .05$) was not significant. In both conditions, as working memory performance (Letter N-Back) increased, VSL-Indirect-LME performance decreased, consistent with the findings in working memory. However, instructional condition did not have an effect on this process.

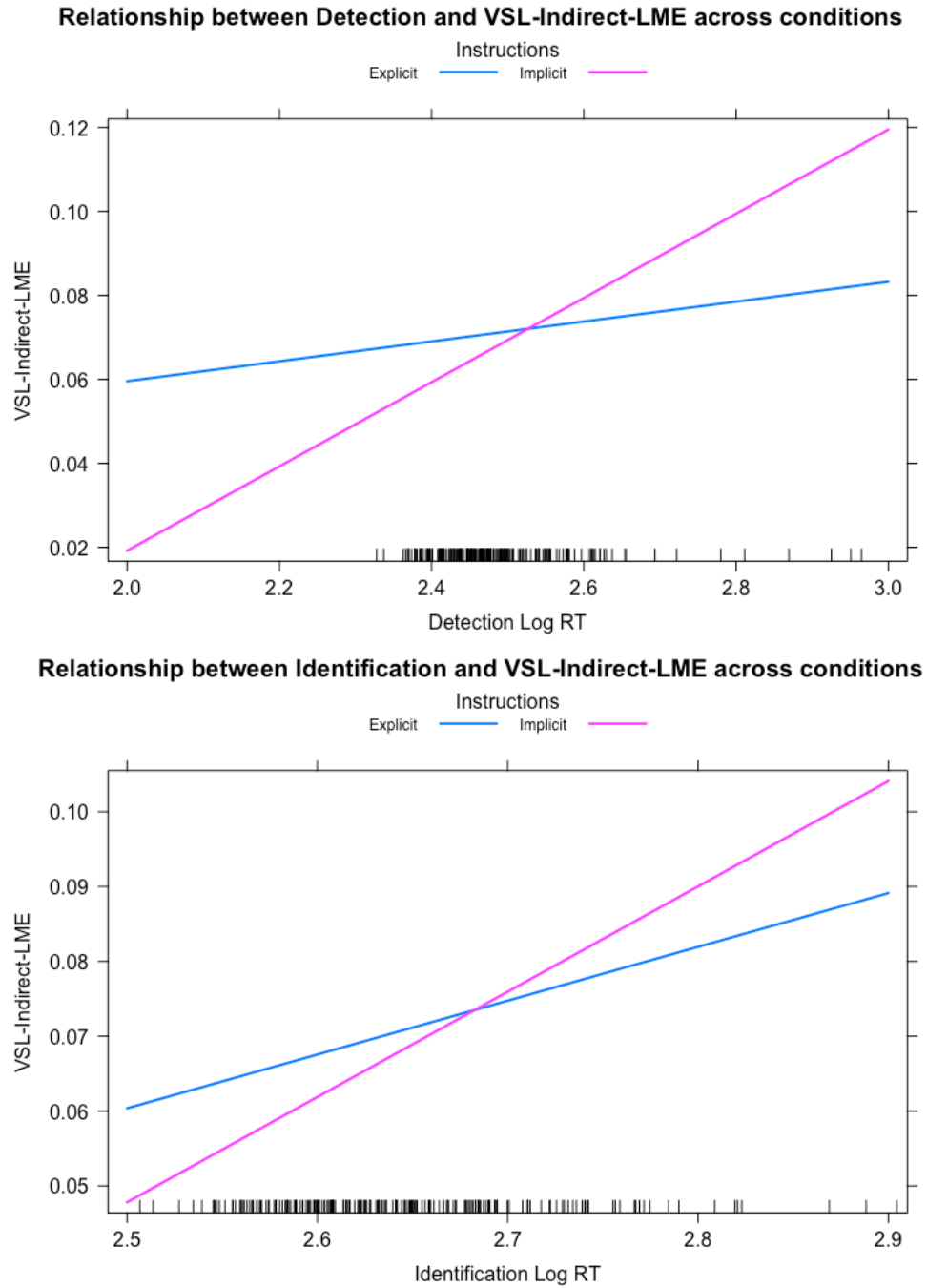


Figure 13. Relationship between VSL-Indirect LME and detection across the Instructional Manipulation.

New Finding: VSL-Indirect was predicted by AGL-clust differently across Instructional conditions

In Model 12.1, AGL-clust was regressed onto VSL-Indirect-LME. AGL-clust marginally significantly predicted VSL-Direct-Logit ($b = .010$, $t(145) = 1.897$, $p = .059$). The interaction between AGL-clust and instructional condition (Model 12.3) significantly improved the fit of the model ($F(2, 143) = 3.60$, $p < .05$). To confirm that the interaction between AGL-clust and instructional condition was the source of the improvement in the model an intermediate model (Model 3.2) was created. Instructional condition (Model 3.2) did not improve the fit of the model over Model 3.1 ($F(1, 144) = 1.25$, $p < .05$). However, including the interaction term improved the fit of the model over Model 3.2 ($F(1, 143) = 5.95$, $p < .05$). In addition, in Model 3.3, the interaction between AGL-clust and instructional condition was significant in the model ($b = .010$, $t(143) = -2.43$, $p < .05$). Interestingly, in Model 3.3, after controlling for instructional condition and the interaction term, AGL-clust was highly significant ($b = .02$, $t(143) = 3.10$, $p < .01$). This suggested across participants, as AGL-clust increased, VSL-Indirect-LME increased. Importantly, the interaction term provides a more complete picture. In the explicit condition, as AGL-clust increased, VSL-Indirect-LME increased. However, in the implicit condition there was no relationship between AGL-clust and VSL-Indirect-LME.

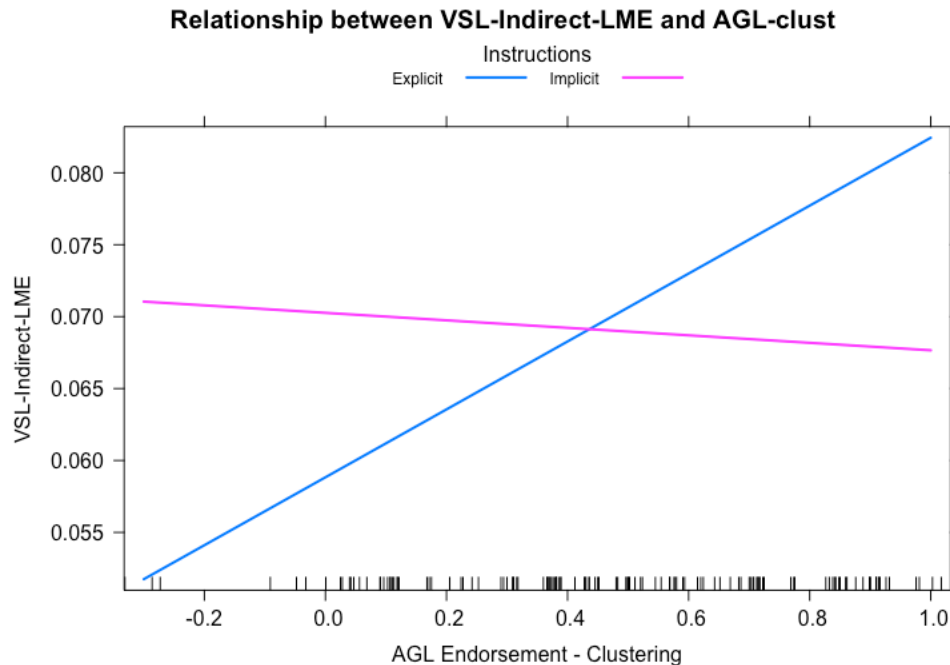


Figure 14. Relationship between VSL-Indirect LME and AGL Endorsement – Clustering across the Instructional Manipulation.

Exploratory Aim 4 Conclusion/Discussion

In Exploratory Aim 4, I re-examined the findings from Aims 1 and 2. Using the updated measures developed in Exploratory Aim 3 both confirmed several of the findings in Aims 1 and 2 and provided additional insights that further develop the theoretical relationship between SL and multiple memory systems. In addition, two marginally significant findings from Aims 1 and 2 became non-significant using the more advanced methods suggesting they were in fact spurious pointing to the strength of the improvements developed in Exploratory Aim 3.

- 1. The updated measures confirmed several findings regarding VSL and the relationship between VSL and IPM and EDM from Aims 1 and 2.**

The results in Aim 4 were largely consistent with Aim 1 and 2. VSL performance on both VSL-Direct-Logit and VSL-Indirect-LME was not affected by the instructional condition and performance on VSL-Direct-Logit and VSL-Indirect-LME were not correlated. In addition, VSL-Direct-Logit was predicted by both paired associate learning (EDM) and SRT (IPM) in both Aim 2 and Aim 4. VSL-Direct-Logit was also predicted by identification (EF) and vocabulary. Turning to VSL-Indirect-LME, the relationship between VSL-Indirect-LME and detection (EF) was confirmed in Aim 4.

2. The updated measures extended the findings regarding VSL and the relationship between VSL and IPM and EDM from Aims 1 and 2 in a manner consistent with the established patterns.

The pattern of results from Aim 4 expanded on previous findings and were consistent with the findings in Aims 1 and 2. For example, the marginally significant relationship between VSL-Indirect and visual object learning-immediate became significant using the updated methods (VSL-Indirect-LME). In addition, consistent with this pattern, visual object learning-delayed (EDM) also became significant using VSL-Indirect-LME. The pattern of the interaction with EDM measures is consistent with the pattern found with executive function. For example, across instructional groups, EDM and executive function skill actually interferes with VSL-Indirect-LME performance (e.g., negative correlations with visual object learning, Detection). However, with EDM, this effect is mostly driven by the participants in the implicit condition. In the explicit condition, EDM has a slight positive-to-null correlation with VSL-Indirect. Whereas in the implicit

condition, as EDM skill increases, VSL-Indirect performance decreases. This pattern is replicated in some of the EF measures in which the interference-type effect of EF on VSL-Indirect performance is mitigated by the explicit instruction set.

3. AGL-clust measures a separate aspect of learning AGL than grammaticality and AGL-clust is more related to EDM processes than IPM.

The pattern of correlation with IPM and EDM measures and the relationship with VSL suggested AGL-clust indexes a separate aspect of learning in AGL than grammaticality that is related to EDM. AGL-clust was negatively correlated with both AGL-d' and AGL-gram. These are measures of sensitivity/memory to/for the more global structure of the input stream, whereas AGL-clust focuses on sensitivity to more local co-occurrences (bigrams, trigrams). This negative correlation suggests that individuals differ in sensitivity to various grain sizes of information and that these forms of information may interfere with each other.

It is interesting to note that the pattern of results for AGL-clust were contrary to what one might predict if the measure is in fact an IPM measure. AGL-clust seems to pattern more like an EDM measure. First, AGL-clust is significantly correlated with both paired associate learning and visual object learning, both EDM measures. Taken by themselves, these correlations may be unconvincing as the overall magnitude of these correlations was quite low. However, like paired associate learning and Identification, measures of EDM and EF respectively, performance on AGL-clust predicted performance

on VSL-Direct-Logit performance across instructional conditions. Additionally, AGL-clust was correlated with vocabulary, a measure related to both VSL-Direct-Logit and EDM. Furthermore, in the implicit condition, AGL-clust does not predict VSL-Indirect-LME as one might expect if AGL-clust was strictly an IPM measure. In fact, in the explicit condition, AGL-clust has a strong positive association with VSL-Indirect-LME performance. Interestingly, this pattern is more in line with initial predictions regarding the effect of instructional condition on the relationship between EDM and VSL than the actual EDM measures (no effect in explicit, interferences in implicit). It is possible that in the explicit condition as participants are told to look for a pattern, they look for or better able to retain more local co-occurrences. Therefore, in the explicit condition, sensitivity to clusters of bigrams and trigrams within an input signal improves overall performance on the indirect measure of VSL.

4. Aim 4 provided additional support to the finding that the instructional condition did not have an effect on the processes underlying VSL-Direct-Logit (Explicit), but VSL-Indirect-LME (Implicit) was affected by the instructional condition.

The pattern of results from Aim 2 were confirmed in Aim 4. Taken together, it is clear that instructional condition shifts processing in the VSL-Indirect-LME task which is an indirect (implicit) measure of VSL, but not in the VSL-Direct-Logit, a more direct (explicit) measure. Overall, VSL-Direct-Logit is related to AGL-clust, paired associate learning, and Identification equally across instructional conditions. Conversely, the

instructional condition affects the relationship between VSL-Indirect-LME and AGL-clust, visual object learning (immediate, delayed), and Detection. Further, VSL-Direct-Logit and VSL-Indirect pattern differently with EDM, IPM and the cognitive measures. For example, examining the relationship between these measures and EDM reveals that EDM supports VSL-Direct performance across conditions. However, EDM has a minimal impact on VSL-Indirect in the explicit condition, but a detrimental impact in the implicit condition. This finding may support the claim that VSL-Direct is a measure more related to EDM than VSL-Indirect (Batterink et al., 2015). Similarly, EF/WM actually supports VSL-Direct, but is detrimental to VSL-Indirect performance. This effect increased in the implicit condition.

General Discussion

The overarching aim of the current study was to characterize the componential nature of SL by understanding its relationship to multiple memory systems (see Fig. 15 for overview). This general aim was addressed with several strategies to approach the problem from multiple angles and obtain a more complete picture. First, as is most common in SL literature (e.g., Arciuli et al., 2014; Yang and Li, 2012), the instructions prior to learning was manipulated to examine differences in performance. However, an underlying assumption within many of these studies is that increased performance with explicit instructions suggests the instruction-type causes additional recruitment of EDM networks. This increased EDM recruitment is then the source of the improvement in performance. Inherent to this assumption is idea that SL is typically driven by IPM processes (see Saffran et al., 1997; Aslin and Newport, 2004; Conway and Christiansen, 2006; Perruchet and Pacton, 2006) and only under certain circumstances does EDM become involved. Evidence from brain data suggests that across modalities and even

within tasks with implicit instructions regions related to both EDM and IPM are activated (e.g., Yang and Li, 2012, see Frost et al., 2015 for review). Furthermore, even in experiments in which there is no difference in performance due to instructions, there were differences in the processes underlying learning (e.g., Yang and Li, 2012; Dienes et al., 1991; Morgan-Short et al., 2012).

While changes in overall performance might provide an initial indication that EDM and IPM are involved in SL, this functions as more of a proof of concept rather than an in-depth exploration of the nature of the relationship. It should be noted that the findings regarding task-dependent and developmental differences that change whether or not there is a difference in performance (e.g., differences found with simple, but not complex patterns; older, but not younger children) may provide additional context to the relationship between SL and multiple memory systems. However, this does not directly examine the nuances in the relationship. Therefore, similar to studies which examined differences in processing underlying performance in explicit and implicit conditions (e.g., Yang and Li, 2012), the current study additionally examined the relationship between individual differences in VSL and EDM, and IPM and cognitive performance. This provides more specific information as to the nuances of the relationship. To provide further context, both indirect and direct measures of VSL were used. These measures have been found to be related to IPM and EDM respectively (Batterink et al., 2015), but the effect of instruction on their relationship to the respective memory systems has not been examined. The connection of language processing, a separate componential cognitive mechanism, to VSL and MMS was also used to confirm this patterning. Furthermore, examination of the relationship between VSL and multiple memory systems

was then used to better understand the widely reported connection between VSL and language as posited by Sawi and Rueckl (2018). The relationship between VSL performance vocabulary was used as a test case to examine some of the predictions generated.

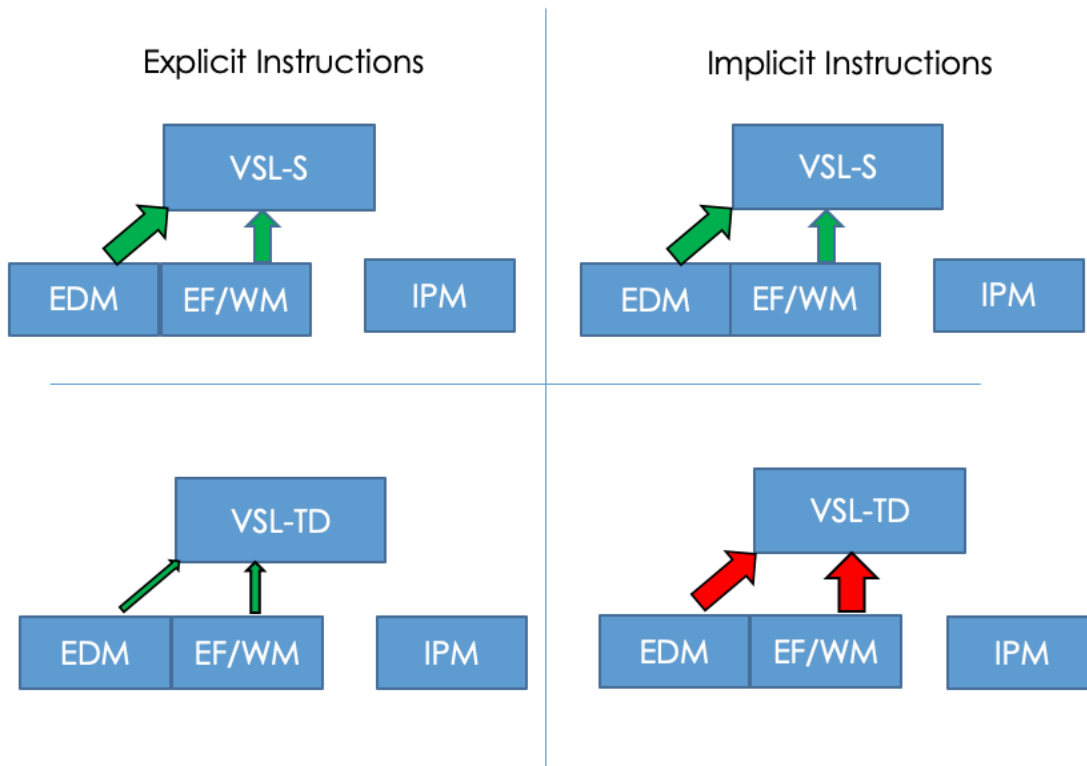


Figure 15. Relationship between VSL and EDM, EF/WM, and IPM. Explicit instructions on the left, Implicit instructions on the right. Direct (explicit) measures on the top row, indirect (implicit) measures on the bottom row. Green arrows represent a positive relationship, red arrows represent a negative relationship. Arrow width represents the strength of the connection.

The Relationship between VSL and Multiple Memory Systems Using Established Measures

Part 1 (Aims 1 and 2) addressed the general aim of the study by using well-established individual differences measures of VSL, EDM, IPM, and general cognition. While there are documented methodological and theoretical issues with some of these measures (VSL and IPM in particular), Part 1 allowed for comparison to findings in the literature. In addition, these established measures created a baseline from which to

compare findings from the updated measures (e.g., if the results were completely different, there may be some issues with the new measures).

The main finding from Part 1 was that VSL-Direct was supported by EDM across conditions, even though there was no difference in performance. This is contrary to many conceptions of SL as strictly IPM. Conversely, VSL-Indirect performance was hindered by executive function. This still implies that VSL-Indirect is related to executive function albeit in a competitive manner. As would be predicted, as EDM and executive function are related constructs (DeKeyser et al., 2003), EDM and executive function support the direct measure of VSL while hindering the indirect measure.

In addition, Part 1 established the pattern in which instructional condition does not significantly impact the processes underlying VSL-Direct but does impact VSL-Indirect. Further, VSL-Direct is supported by EDM across conditions and paired associate learning in particular. This is consistent with the fact that MTL is active regardless of instructional condition (e.g., Yang and Li, 2012) and paired associate learning is strongly dependent on the MTL network (see Krishnan et al., 2016; Suzuki, 2008 for review). In addition, VSL-Indirect performance is inhibited by executive function and this effect is increased in the implicit condition possibly suggesting a disruption of the direct cortical-subcortical activation pattern typically found in implicit learning by increased activation from attentional networks (Yang and Li, 2012) or the competitive nature of EDM and IPM processes in learning (Poldrack et al., 2001). Further confirming the pattern described previously, VSL-Direct performance positively predicted vocabulary across both instruction groups and this relationship did not shift due to the instruction type. In addition, only the direct (VSL-Direct) and not the indirect (VSL-Indirect) measure of VSL predicted

performance. Taken together with the previous results, this suggests that VSL-Direct and EDM support similar aspects of language processing and are related constructs.

If only the effect on performance was examined, much of the nuanced exploration of the relationship between SL and multiple memory systems would be lost as there were no differences across any of the VSL task. As the experiment currently stands, Part 1 provided insight into the relationship between VSL and EDM in particular. However, the relationship between these constructs has interesting nuances, rather than simply stating VSL involves EDM and EF (e.g., differential directions of association and effect of instructions across measurement type). Manipulating both instructional condition and measurement type allowed for the discoveries described above to be better understood.

Optimizing Methods for VSL and IPM Individual Differences Analyses

Part 2 (Exploratory Aim 3) addressed some of the methodological and theoretical issues with measurement in SL and IPM. Addressing these issues allowed for the expansion of the findings from Part 1. However, as these measures are not as well established in the literature and represent an extension of current measurement techniques, Part 2 was more exploratory. First, it should be noted that in Part 1, both VSL-SPT (self-paced task) and Categorization participants did not show significant learning. An important finding regarding VSL-SPT, in line with the goals of Aim 3, is that VSL-SPT is greatly impacted by the use of a repeated item cover task (as found in Arciuli et al., 2012). As VSL-SPT is a relatively new measure, appearing in only two publications from the same group (Siegelman et al., 2019a; Siegelman et al., 2019b), the cover task and instructions are important considerations. The cover task was needed to facilitate participants even being able to complete the task with only implicit instructions (without

guidance, participants may simply press the button to continue very quickly only looking for repeats). More explicit type instructions (and no cover task) may be necessary prerequisites for the VSL Self-paced task to provide an accurate learning measure.

Returning to Part 2, Part 2 included several important take-aways for developing more accurate measures of SL and IPM. First, in some cases and particularly with accuracy, sometimes the simplest method that still addresses the methodological and theoretical issues is the most effective (e.g., AGL-d', VSL-Direct Logit transformation). In most cases, the fewer assumptions about the data that need to be made and the fewer algorithms and/or transformations that need to be used to provide a clean signal, the better. Furthermore, across the board, LME coefficients provided the most reliable measures overall. This is likely due to the fact that LME takes all of the data (group-level effects also). Note, the incorporation of group-level effects may also be a double-edged sword as this process introduces artifacts such as data shrinkage. Therefore, it is important to balance reliability with concerns regarding validity. In addition, another interesting finding is that across the board split-half reliability for individual regression models is very low, often lower than even basic mean scores and only slightly better than mean difference measures. As discussed, this may due to the fact that the data are split, and the signal-to-noise ratio may just be artificially deflated causing variability in responses. Future studies should explore this issue using test-retest reliability and/or using many more trials so the impact of the split is lessened.

In addition, in creating individual regression models for use in individual differences analyses, an important first step is to generate an LME model. Specifically, the random effects structure helps define which variables actually vary at an individual level. Inclusion

of non-significant random effects generates coefficients with very low reliability (and even lowers the reliability of significant random effects). Relatedly, inclusion of interaction terms in individual regressions generated measures with incredibly low reliability (close to zero). This may be due to the fact that none of the interactions were significant at the random-effects level, but the interaction terms were even less reliable than regular non-significant predictors.

Lastly, Part 2 addressed methodological and theoretical issues with the version of the AGL used in Part 1 (outcome measure: AGL proportion correct). Namely, that as the measure is currently defined, AGL more likely measures familiarity of more global-level co-occurrence generated through IPM processes (but not directly if participants learned the “grammar” or are making “grammaticality judgment”). Furthermore, and importantly for Aim 4, “grammaticality” was fully crossed with clustering (frequency of bigrams and trigrams present in each item) in order to control for the influence of bigram/trigram frequency. As the task is currently designed, sensitivity to clusters was occluded by using the “grammaticality” judgment and in turn clustering affected performance on “grammaticality”. Endorsement rate was used instead of accuracy to shift focus to the sensitivity to these factors that measure different grain sizes of information. LME generated cleaner measures of grammaticality and clustering. Interestingly, these factors have an inhibitory/competitive relationship. Furthermore, as found in Aim 4, they pattern differently with AGL-clust seemingly more akin to EDM than IPM.

Exploring the Relationship between VSL and multiple memory Multiple Memory Systems

Part 3 (Exploratory Aim 4) used the updated measures from Part 2 and expanded the findings from Part 1. It is important to note that all but one finding from Part 1 was confirmed in Part 3 (marginal interaction between VSL-Indirect and interaction on the effect of VSL-Direct). The exploratory updated measures provided additional insights consistent with the patterns established in Part 1.

For example, in Part 1, VSL-Direct was supported by EDM. The expanded findings from Part 3 are consistent with this result as VSL-Direct-Logit was supported by executive function and AGL-clust, two constructs related to EDM. In addition, the relationship between VSL and EDM, EF, and AGL-clust were not affected by instruction. Turning to VSL-Indirect-LME, in Part 1, executive function interfered with VSL-Indirect-LME performance with the interference effect significantly increased in the implicit condition. In Part 3, EDM additionally followed the same pattern as executive function. However, AGL-clust instead positively predicted VSL-Indirect-LME performance in the explicit condition but did not have a relationship with VSL-Indirect-LME in the implicit condition. While this pattern is not exactly the same as with executive function and EDM, it is what would be predicted if the instructional condition shifts processing of the more indirect measure towards EDM.

Taken together, while VSL-Direct-Logit is not affected by instructional condition, VSL-Indirect-LME relies more on EDM in the explicit condition (e.g., interaction with AGL-clust and instructional condition). In the implicit condition on the other hand, the involvement of IPM is enhanced (e.g., interference from EDM and executive function). As was the case with Part 1, VSL-Direct-Logit was predicted by Vocabulary, but VSL-Indirect-LME was not. EDM also predicted vocabulary performance. Additionally, further

linking AGL-clust to aspects of EDM processing, AGL-clust also predicted vocabulary performance.

Characterizing the Componential Nature of SL

The current study sought to further expand understanding of the componential nature of SL. Much work has been done on establishing the modality-specific components of SL (see Frost et al., 2015; Siegelman et al., 2018 for review). With particular focus on specific neurocognitive mechanisms which may drive individual differences. While the authors focus on modality-specific mechanisms driving SL, Frost et al. (2015) also posit potential mechanisms driving domain-general components of SL. Specifically, modality-specific information generated during initial encoding is further processed in multimodal regions (e.g., frontal, striatal and MTL memory systems). Information across all domains is therefore processed in the same brain networks and may be subject to similar processing demands. However, the nature of the processing in these multimodal regions is not described in detail.

There have been several theoretical frameworks which have explored these domain general components. For example, Thiessen and Erickson (2013, 2015) modeled the underlying processes supporting sensitivity to multiple forms of statistical information across modality. In their model, conditional statistics and distributional statistics are modeled by different underlying computational, memory-based systems. However, these systems are also linked, as the output of computations related to conditional statistics (extraction) provide the input for processes involved in the computation of distributional statistics (integration).

Further, Arciuli (2017) proposed that SL is supported by aspects of memory such as encoding and retention and working memory. Additionally, both Arciuli (2017) and Gomez (2017) suggested a developmental time scale for the engagement of specific memory systems such as WM and EDM during SL, suggesting in order to function optimally, SL requires input from a constellation of interrelated cognitive functions related to EDM, IPM, and executive function. However, the arguments for the engagement of multiple memory systems presented in Arciuli (2017) and Gomez (2017) were based on differences in findings across studies with different populations, making direct comparison difficult (speculative). This issue is similar to the examination of the presence or absence of differences in VSL performance due to instructional manipulations across various studies in order to make claims about how and when multiple memory systems support SL (e.g., Arciuli et al., 2015). Grounding the componential nature of the domain-general aspects of SL in theories of multiple memory systems has been particularly powerful in understanding the role of multimodal processing systems (e.g., MTL, frontal-striatal networks). In addition, invoking this particular literature allows one to leverage the understanding of the interactivity of these memory systems (e.g., differences in underlying processing in the absence of differences in performance) to explain SL processing.

A strength of the current study is the simultaneous inclusion of measures for VSL, EDM, IPM, and EF/WM. In addition, to develop a more nuanced understanding of the relationship between these measures both instruction and measurement type were manipulated. Inclusion of all of these measures and manipulations allows for a more direct comparison of the relationship and exploration of exactly how and when they interact. Following a line of research suggesting SL is related to EDM (e.g., Kachergis et al., 2010;

Hamrick and Rebuschat, 2012) contrary to many early conceptions of SL as purely IPM (e.g., Conway and Christiansen, 2006; Perruchet and Pacton, 2006), the current study found that VSL was related to EDM and executive function. It is interesting to note, much of the early work with SL, which assumed a purely IPM basis, used a measure equivalent to the VSL-Direct, the direct measure of SL. Prior research found this measure to be related to ERPs indicative EDM processes (Batterink et al., 2015). The current study found that VSL is supported by some aspects of EDM and EF. In both instructional conditions VSL-Direct was supported to half of the measures of EDM and executive function. Even the indirect VSL measure, VSL-Indirect, which is related to IPM (Batterink et al., 2015) was related supported by some, albeit different, aspects of EDM and executive function.

However, there were interesting nuances in this relationship. The direct, post-learning measure did not change underlying EDM processes significantly due to instructions. It is possible that VSL-Direct does in fact shift processing, but additional brain data would need to be collected. For example, Schenden et al. (2003) posited that MTL is involved in sequence learning in EDM and IPM but engage anterior and posterior regions respectively. VSL-Indirect, the indirect semi-learning measure, does in fact shift due to instructions. VSL-Indirect was correlated with AGL-clust, but only in the explicit condition. This is potentially indicative of the additional activation of EDM regions such as the precuneus or attentional networks found during SL with explicit instructions (Yang and Li, 2012). As stated previously, the interference effect of EDM and EF/WM found in the implicit condition might indicate a disruption of the direct cortical-subcortical connection

typically found in SL with implicit conditions (Yang and Li, 2012) or the overall competitive nature of learning in EDM and IPM (Poldrack et al., 2001).

In conclusion, it is clear that aspects of IPM, EDM, and executive function underlie domain-general components of SL. The nature of this relationship is shifted in interesting ways due to specific task demands (instructions) and measurement (direct, indirect) reflecting differential engagement of a set of interrelated neurocognitive mechanisms related to MMS. Exploring SL in this manner provides additional context for the interactivity in the engagement of these multimodal processing centers as described in Frost et al. (2015). Insights from the MMS literature directed manipulation of SL (instructions, measurement) and provided a framework for understanding the results (e.g., competitive nature of MMS engagement, Poldrack et al., 2001).

Understanding the Relationship between VSL and Language

The findings from the overarching aim of the study may be used to provide context for an important current issue within SL literature: the connection between SL and language processing. The relationship between VSL and vocabulary can be used as an interesting test case for use of the expanded understanding of the componential nature of VSL.

VSL-Direct performance positively predicted vocabulary performance, such that higher VSL-Direct performance was associated with individuals with higher vocabulary scores across both instruction groups, consistent with recent evidence (Sawi and Rueckl, 2018) pointing to a link between SL and language processing. In addition, only the direct (VSL-Direct) and not the indirect (VSL-Indirect) measure of VSL predicted performance. This suggests that the VSL may be related to vocabulary through more EDM rather than

IPM processes. These types of results are not novel and have been reported in many previous studies with children and adults (e.g., Shafto et al., 2012; Mainela-Arnold and Evans, 2014). However, these studies typically do not provide additional context as the specific mechanisms underlying this relationship. The question addressed by these studies is often simplified to “Does statistical learning ability predict language?” However, neither language processing nor SL are monolithic skills. Rather, both are fundamentally componential requiring the engagement of distinct neurocognitive networks (see Sawi and Rueckl, 2018 for a review). Here we may provide an initial exploratory step in extending the literature by asking *how* SL and language are related rather than just *if*. For example, several frameworks have posited that the MTL plays a critical role in the learning of arbitrary associations (McClelland et al., 1995; Squire, 1992). This suggests that, like with paired associate learning, MTL mediation would be particularly important in learning along the semantic pathway. Consistent with this assertion, vocabulary seems to be related to the more explicit aspects of VSL (VSL-Direct) controlling for EDM. These findings suggest that VSL-Direct predicts vocabulary through the engagement of the MTL network during the learning of arbitrary relationships. Furthermore, individuals with greater reliance on semantic information during reading may have greater MTL network activation relative to frontal-striatal networks.

Conclusion

Statistical Learning is a deeply componential construct. Understanding the componential nature of SL is of paramount importance in order to understand the relationship between SL and other cognitive systems such as reading and memory. The current study sought to expand the literature by exploring the relationship between SL

and multiple memory systems. SL shares several neural correlates with aspects of EDM and IPM and performance on SL can be manipulated by memory-related factors (e.g., instruction-type, age). In addition, grounding SL in multiple memory system theories can help provide additional insight into componentiality in SL as memory is also a deeply componential construct.

Exploring the connection to multiple memory systems theory yielded several important findings. For example, VSL performance (direct, indirect) was not affected by the instructional manipulation. This is contrary to the assumption in the field that SL is typically driven by IPM processes and only under certain circumstances does EDM become involved (see Saffran et al., 1997; Aslin and Newport, 2004; Conway and Christiansen, 2006; Perruchet and Pacton, 2006). This may suggest both EDM and IPM are active during SL as in Yang and Li (2012). Additionally, with no change in performance, evidence suggests that the instructional manipulation shifted the processes underlying VSL. While VSL-Direct was not affected by the instructional manipulation, VSL-Indirect seemed to differentially engage EDM/EF (greater EDM engagement in explicit condition, greater IPM engagement in the implicit condition). Furthermore, while there were inconsistencies in the results, patterning suggests VSL may be related to aspects of both EDM and IPM. This is consistent with recent data suggesting both EDM and IPM support VSL (e.g., Batterink et al., 2015; Kachergis et al., 2010; Hamrick and Rebuschat, 2012). In addition, evidence suggests there is domain general activation of both MTL and Frontal-striatal memory networks (EDM, IPM) (Frost et al., 2015) across SL tasks.

However, overall findings were inconsistent. For example, EDM and IPM measures were not correlated. In addition, there were inconsistencies within each construct (i.e., EDM measures were not correlated with each other). In addition, VSL-Direct and VSL-Indirect were related to different sets of EDM/EF measures. These inconsistencies may point to componentiality in VSL, EDM and IPM. However, this points to deep methodological concerns (e.g., these measures are not properly optimized for individual differences analyses). For future studies, it will be important to develop between measures of VSL and IPM in particular. Specifically, it is important to develop more reliable versions of these measures (Siegelman et al., 2015) that balance validity, reliability, and simplicity of approach. However, it is important to note that statistically controlling for specific nuisance variables does Not constitute a “cure-all” for these methodological concerns. More in-depth analyses of the psychometric properties of VSL and IPM measures are needed.

References

- Alba, D., and Okanoya, K. (2008). Statistical segmentation of tone sequences activates the left inferior frontal cortex: A near-infrared spectroscopy study. *Neuropsychologia*, 46(11), 2787–2795.
<https://doi.org/10.1016/j.neuropsychologia.2008.05.012>
- Apfelbaum, K. S., Hazeltine, E., and McMurray, B. (2013). Statistical learning in reading: Variability in irrelevant letters help children learn phonics skills. *Developmental Psychology*, 49(7), 1348–1365. doi:10.1037/a0029839
- Arciuli, J. (2017). The multi-component nature of statistical learning. *Phil. Trans. R. Soc. B*, 372(1711), 20160058. doi:10.1098/rstb.2016.0058.
- Arciuli, J., and Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, 14(3), 464–473. doi:10.3389/fpsyg.2014.01541
- Arciuli, J., and Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience letters*, 517(2), 133–135. doi:10.1016/j.neulet.2012.04.045
- Arciuli, J., Torkildsen, J. V. K., Stevens, D. J., and Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in psychology*, 5, 747. doi:10.3389/fpsyg.2014.00747
- Aslin, R. N., and Newport, E. L. (2012). Statistical Learning. *Current Directions in Psychological Science*, 21(3), 170–176. doi:10.1177/0963721412436806
- Baayen, R. H., and Ramscar, M. (2015). Abstraction, storage and naive discriminative learning. In E. Dabrowska and D. Divjak (Eds.), *Handbook of cognitive linguistics* (Vol. 39) (pp. 99–120). Berlin: Walter de Gruyter GmbH and Co KG.
- Baron, J., and Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 2(3), 386–393. doi:10.1037//0096-1523.2.3.386
- Batterink, L. J., Reber, P. J., Neville, H. J., and Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, 83, 62–78. doi:10.1016/j.jml.2015.04.004
- Bertels, J., Boursain, E., Destrebecqz, A., and Gaillard, V. (2015). Visual statistical learning in children and young adults: how implicit? *Frontiers in Psychology*, 5. doi:10.3389/fpsyg.2014.01541
- Bertels, J., Franco, A., and Destrebecqz, A. (2012). How implicit is visual statistical learning?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5), 1425.
- Bischoff-Grethe, A., Proper, S. M., Mao, H., Daniels, K. A., and Berns, G. S. (2000). Conscious and unconscious processing of nonverbal predictability in Wernicke's area. *Journal of Neuroscience*, 20(5), 1975–1981.
- Bogaerts, L., Szmalec, A., Hachmann, W. M., Page, M. P., and Duyck, W. (2015). Linking memory and language: Evidence for a serial-order learning impairment in dyslexia. *Research in Developmental Disabilities*, 43–44, 106–122. doi:10.1016/j.ridd.2015.06.012
- Browman, C. P., and Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6(02), 201–251. doi:10.1017/s0952675700001019

- Brown, J., Aczel, B., Jiménez, L., Kaufman, S. B., and Grant, K. P. (2010). Intact implicit learning in autism spectrum conditions. *Quarterly Journal of Experimental Psychology*, 63(9), 1789-1812.
- Bussy, G., Krifi-Papoz, S., Vieville, L., Frenay, C., Curie, A., Rousselle, C., Rougeot, C., Des Portes, V., Herbillon, V. (2011). Apprentissage procédural implicite dans la dyslexie de surface et la dyslexie phonologique. *Revue de neuropsychologie*, 3(3), 141. doi:10.3917/rne.033.0141
- Castles, A., and Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47(2), 149-180. doi:10.1016/0010-0277(93)90003-e
- Chang, L., Plaut, D. C., and Perfetti, C. A. (2016). Visual complexity in orthographic learning: Modeling learning across writing system variations. *Scientific Studies of Reading*, 20(1), 64-85. doi:10.1080/10888438.2015.1104688
- Christiansen, M. H., Conway, C. M., and Curtin, S. (2005). Multiple-cue integration in language acquisition: A connectionist model of speech segmentation and rule-like behavior. *Language acquisition, change and emergence: Essay in evolutionary linguistics*, 205-249.
- Christiansen, M. H., Dale, R., and Real, F. (2009). Connectionist Explorations of Multiple-Cue Integration in Syntax Acquisition. *Neoconstructivism*, 87-108. doi:10.1093/acprof:oso/9780195331059.003.0005
- Chun, M. M., and Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, 36(1), 28-71. doi:10.1006/cogp.1998.0681
- Cohen, N. J., Poldrack, R. A., and Eichenbaum, H. (1997). Memory for Items and Memory for Relations in the Procedural/Declarative Memory Framework. *Memory*, 5(1-2), 131-178. doi:10.1080/741941149
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204-256. doi:10.1037//0033-295x.108.1.204
- Conway, C. M., and Christiansen, M. H. (2005). Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 24-39. doi:10.1037/0278-7393.31.1.24
- Conway, C. M., and Christiansen, M. H. (2006). Statistical Learning Within and Between Modalities. *Psychological Science*, 17(10), 905-912. doi:10.1111/j.1467-9280.2006.01801.x
- Conway, C. M., and Pisoni, D. B. (2008). Neurocognitive Basis of Implicit Learning of Sequential Structure and Its Relation to Language Processing. *Annals of the New York Academy of Sciences*, 1145(1), 113-131. doi:10.1196/annals.1416.009
- Conway, C. M., Bauernschmidt, A., Huang, S. S., and Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, 114(3), 356-371. doi:10.1016/j.cognition.2009.10.009
- Davis, M. H., Betta, A. M., Macdonald, M. J., and Gaskell, M. G. (2009). Learning and Consolidation of Novel Spoken Words. *Journal of Cognitive Neuroscience*, 21(4), 803-820. doi:10.1162/jocn.2009.21059
- Dekeyser, R. (2003). Implicit and Explicit Learning. *The Handbook of Second Language Acquisition*, 312-348. doi:10.1002/9780470756492.ch11

- Deroost, N., Zeischka, P., Coomans, D., Bouazza, S., Depessemer, P., and Soetens, E. (2010). Intact first- and second-order implicit sequence learning in secondary-school-aged children with developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, 32(6), 561-572. doi:10.1080/13803390903313556
- Destrebecqz, A. (2005). The neural correlates of implicit and explicit sequence learning: Interacting networks revealed by the process dissociation procedure. *Learning and Memory*, 12(5), 480-490. doi:10.1101/lm.95605
- Destrebecqz, A., and Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin and Review*, 8(2), 343-350. doi:10.3758/bf03196171
- Dew, I. T., and Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: behavioral and neural evidence. *Annals of the New York Academy of Sciences*, 1224(1), 174-190. doi:10.1111/j.1749-6632.2010.05946.x
- Dienes, Z., Broadbent, D., and Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(5), 875-887. doi:10.1037//0278-7393.17.5.875
- Durrant, S. J., Cairney, S. A., and Lewis, P. A. (2012). Overnight Consolidation Aids the Transfer of Statistical Knowledge from the Medial Temporal Lobe to the Striatum. *Cerebral Cortex*, 23(10), 2467-2478. doi:10.1093/cercor/bhs244
- Durrant, S. J., Taylor, C., Cairney, S., and Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, 49(5), 1322-1331. doi:10.1016/j.neuropsychologia.2011.02.015
- Eichenbaum, H. (2001). The hippocampus and declarative memory: cognitive mechanisms and neural codes. *Behavioural Brain Research*, 127(1-2), 199-207. doi:10.1016/s0166-4328(01)00365-5
- Eichenbaum, H., and Cohen, N. J. (2001). From conditioning to conscious recollection: Memory systems of the brain. New York: Oxford University Press.
- Endress, A. D., and Mehler, J. (2009). The surprising power of statistical learning: When fragment knowledge leads to false memories of unheard words. *Journal of Memory and Language*, 60(3), 351-367.
- Erickson, L. C., and Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, 37, 66-108. doi:10.1016/j.dr.2015.05.002
- Fischer, S., Drosopoulos, S., Tsen, J., and Born, J. (2006). Implicit Learning—Explicit Knowing: A Role for Sleep in Memory System Interaction. *Journal of Cognitive Neuroscience*, 18(3), 311-319. doi:10.1162/089892906775990598
- Fiser, J., and Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458-467. doi:10.1037//0278-7393.28.3.458
- Folia, V., Uddén, J., Forkstam, C., Ingvar, M., Hagoort, P., and Petersson, K. M. (2008). Implicit Learning and Dyslexia. *Annals of the New York Academy of Sciences*, 1145(1), 132-150. doi:10.1196/annals.1416.012
- Frensch, P. A., and Miner, C. S. (1994). Effects of presentation rate and individual differences in short-term memory capacity on an indirect measure of serial learning. *Memory and Cognition*, 22(1), 95-110. doi:10.3758/bf03202765

- Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(05), 263-279. doi:10.1017/s0140525x11001841
- Frost, R., Armstrong, B. C., Siegelman, N., and Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in Cognitive Sciences*, 19(3), 117-125. doi:10.1016/j.tics.2014.12.010
- Frost, R., Katz, L., and Bentin, S. (1987). Strategies for visual word recognition and orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance*, 13(1), 104-115. doi:10.1037//0096-1523.13.1.104
- Frost, R., Siegelman, N., Narkiss, A., and Afek, L. (2013). What Predicts Successful Literacy Acquisition in a Second Language? *Psychological Science*, 24(7), 1243-1252. doi:10.1177/0956797612472207
- Gabay, Y., Schiff, R., and Vakil, E. (2012). Dissociation between online and offline learning in developmental dyslexia. *Journal of Clinical and Experimental Neuropsychology*, 34(3), 279-288. doi:10.1080/13803395.2011.633499
- Gebhart, A. L., Newport, E. L., and Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin and Review*, 16(3), 486-490. doi:10.3758/pbr.16.3.486
- Goldman-Rakic, P., Selemon, L., and Schwartz, M. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience*, 12(3), 719-743. doi:10.1016/0306-4522(84)90166-0
- Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160054. doi:10.1098/rstb.2016.0054
- Gómez, R. L., and Edgin, J. O. (2016). The extended trajectory of hippocampal development: Implications for early memory development and disorder. *Developmental Cognitive Neuroscience*, 18, 57-69. doi:10.1016/j.dcn.2015.08.009
- Gomez, R. L., and Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, 70(2), 109-135.
- Goujon, A., Didierjean, A., and Thorpe, S. (2015). Investigating implicit statistical learning mechanisms through contextual cueing. *Trends in Cognitive Sciences*, 19(9), 524-533. doi:10.1016/j.tics.2015.07.009
- Graves, W. W., Desai, R., Humphries, C., Seidenberg, M. S., and Binder, J. R. (2010). Neural Systems for Reading Aloud: A Multiparametric Approach. *Cerebral Cortex*, 20(8), 1799-1815. doi:10.1093/cercor/bhp245
- Hachmann, W. M., Bogaerts, L., Szmalec, A., Woumans, E., Duyck, W., and Job, R. (2014). Short-term memory for order but not for item information is impaired in developmental dyslexia. *Annals of dyslexia*, 64(2), 121-136.
- Hamrick, P., and Rebuschat, P. (2012) How implicit is statistical learning? *Statistical Learning and Language Acquisition*. doi:10.1515/9781934078242.365
- Harm, M. W., and Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 491-528. doi:10.1037//0033-295x.106.3.491

- Harm, M. W., and Seidenberg, M. S. (2004). Computing the Meanings of Words in Reading: Cooperative Division of Labor Between Visual and Phonological Processes. *Psychological Review*, 111(3), 662-720. doi:10.1037/0033-295x.111.3.662
- Henderson, L. M., and Warmington, M. (2017). A sequence learning impairment in dyslexia? It depends on the task. *Research in Developmental Disabilities*, 60, 198-210. doi:10.1016/j.ridd.2016.11.002
- Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393-402.
- Howard, J. H., Howard, D. V., Japikse, K. C., and Eden, G. F. (2006). Dyslexics are impaired on implicit higher-order sequence learning, but not on implicit spatial context learning. *Neuropsychologia*, 44(7), 1131-1144. doi:10.1016/j.neuropsychologia.2005.10.015
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of memory and language*, 30(5), 513-541.
- Jeste, S. S., Kirkham, N., Senturk, D., Hasenstab, K., Sugar, C., Kupelian, C., Baker, E., Sanders, A. J., Shimizu, C., Norona, A., Paparella, T., Freeman, S. F. N., Johnson, S. P. (2015). Electrophysiological evidence of heterogeneity in visual statistical learning in young children with ASD. *Developmental Science*, 18(1), 90-105. doi:10.1111/desc.12188
- Jiménez-Fernández, G., Vaquero, J. M., Jiménez, L., and Defior, S. (2010). Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. *Annals of dyslexia*, 61(1), 85-110.
- Jiménez, L., Méndez, C., and Cleeremans, A. (1996). Comparing direct and indirect measures of sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(4), 948-969. doi:10.1037//0278-7393.22.4.948
- Karza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., and Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language*, 127(1), 46-54. doi:10.1016/j.bandl.2012.11.007
- Kelly, S. W., Griffiths, S., and Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, 8(1), 43-52.
- Kidd, E., and Arciuli, J. (2016). Individual Differences in Statistical Learning Predict Childrens Comprehension of Syntax. *Child Development*, 87(1), 184-193. doi:10.1111/cdev.12461
- Kim, R., Seitz, A., Feenstra, H., and Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, 461(2), 145-149. doi:10.1016/j.neulet.2009.06.030
- Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83(2). doi:10.1016/s0010-0277(02)00004-5
- Lerner, I., Armstrong, B. C., and Frost, R. (2014). What can we learn from learning models about sensitivity to letter-order in visual word recognition? *Journal of Memory and Language*, 77, 40-58. doi:10.1016/j.jml.2014.09.002
- Lieberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1), 1-36. doi:10.1016/0010-0277(85)90021-6

- Liu, G., Yang, J., and Li, Z. (2015). Content-based image retrieval using computational visual attention model. *Pattern Recognition*, 48(8), 2554-2566. doi:10.1016/j.patcog.2015.02.005
- Lum, J. A., Ullman, M. T., and Conti-Ramsden, G. (2013). Procedural learning is impaired in dyslexia: Evidence from a meta-analysis of serial reaction time studies. *Research in Developmental Disabilities*, 34(10), 3460-3476. doi:10.1016/j.ridd.2013.07.017
- Mainela-Arnold, E., and Evans, J. L. (2014). Do statistical segmentation abilities predict lexical-phonological and lexical-semantic abilities in children with and without SLI? *Journal of Child Language*, 41(02), 327-351. doi:10.1017/s0305000912000736
- Marshall, L., and Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences*, 11(10), 442-450. doi:10.1016/j.tics.2007.09.001
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3). doi:10.1016/s0010-0277(01)00157-3
- Mayo, J., and Eigsti, I. M. (2012). Brief report: A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of autism and developmental disorders*, 42(11), 2476-2485.
- McClelland, J. L., McNaughton, B. L., and Oreilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3), 419-457. doi:10.1037//0033-295x.102.3.419
- Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., Ruffino, M., Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: A comparative study. *Neuropsychologia*, 48(4), 863-872. doi:10.1016/j.neuropsychologia.2009.11.003
- Menghini, D., Hagberg, G. E., Caltagirone, C., Petrosini, L., and Vicari, S. (2006). Implicit learning deficits in dyslexic adults: An fMRI study. *NeuroImage*, 33(4), 1218-1226. doi:10.1016/j.neuroimage.2006.08.024
- Menghini, D., Hagberg, G. E., Petrosini, L., Bozzali, M., Macaluso, E., Caltagirone, C., and Vicari, S. (2008). Structural Correlates of Implicit Learning Deficits in Subjects with Developmental Dyslexia. *Annals of the New York Academy of Sciences*, 1145(1), 212-221. doi:10.1196/annals.1416.010
- Misyak, J. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1. doi:10.3389/fpsyg.2010.00031
- Misyak, J. B., and Christiansen, M. H. (2012). Statistical Learning and Language: An Individual Differences Study. *Language Learning*, 62(1), 302-331. doi:10.1111/j.1467-9922.2010.00626.x
- Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., Katz, L., Francis, D. J., Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology*, 90(3), 347-373. doi:10.1037//0022-0663.90.3.347

- Newport, E. L., and Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive psychology*, 48(2), 127-162.
- Nigro, L., Jiménez-Fernández, G., Simpson, I. C., and Defior, S. (2015). Implicit Learning of Written Regularities and Its Relation to Literacy Acquisition in a Shallow Orthography. *Journal of Psycholinguistic Research*, 44(5), 571-585. doi:10.1007/s10936-014-9303-9
- Nissen, M. J.; Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*. 19: 1. doi:10.1016/0010-0285(87)90002-8.
- Packard, M. G., and Knowlton, B. J. (2002). Learning and Memory Functions of the Basal Ganglia. *Annual Review of Neuroscience*, 25(1), 563-593. doi:10.1146/annurev.neuro.25.112701.142937
- Page, M. P., Cumming, N., Norris, D., Hitch, G. J., and Mcneil, A. M. (2006). Repetition learning in the immediate serial recall of visual and auditory materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 716-733. doi:10.1037/0278-7393.32.4.716
- Perruchet, P., and Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10(5), 233-238. doi:10.1016/j.tics.2006.03.006
- Petersson, K., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, 120(2), 83-95. doi:10.1016/j.bandl.2010.08.003
- Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Moyano, J. C., Myers, C., and Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature*, 414(6863), 546-550. doi:10.1038/35107080
- Pugh, K. R., Mencl, W. E., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Constable, R. T., Skudlarski, P., Marchione, K. E., Jenner, A. R., Fletcher, J. M., Liberman, A. M., Shankweiler, D. P., Katz, L., Lacadi, C., Gore, J. C. (2000). The Angular Gyrus in Developmental Dyslexia: Task-Specific Differences in Functional Connectivity Within Posterior Cortex. *Psychological Science*, 11(1), 51-56. doi:10.1111/1467-9280.00214
- Reber, A. S. (1992). The cognitive unconscious: An evolutionary perspective. *Consciousness and cognition*, 1(2), 93-133.
- Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia*, 51(10), 2026-2042. doi:10.1016/j.neuropsychologia.2013.06.019
- Reber, P. J., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2003). Dissociating Explicit and Implicit Category Knowledge with fMRI. *Journal of Cognitive Neuroscience*, 15(4), 574-583. doi:10.1162/089892903321662958
- Redington, M., and Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of experimental psychology: general*, 125(2), 123.
- Robertson, E. M. (2007). The Serial Reaction Time Task: Implicit Motor Skill Learning? *Journal of Neuroscience*, 27(38), 10073-10075. doi:10.1523/jneurosci.2747-07.2007

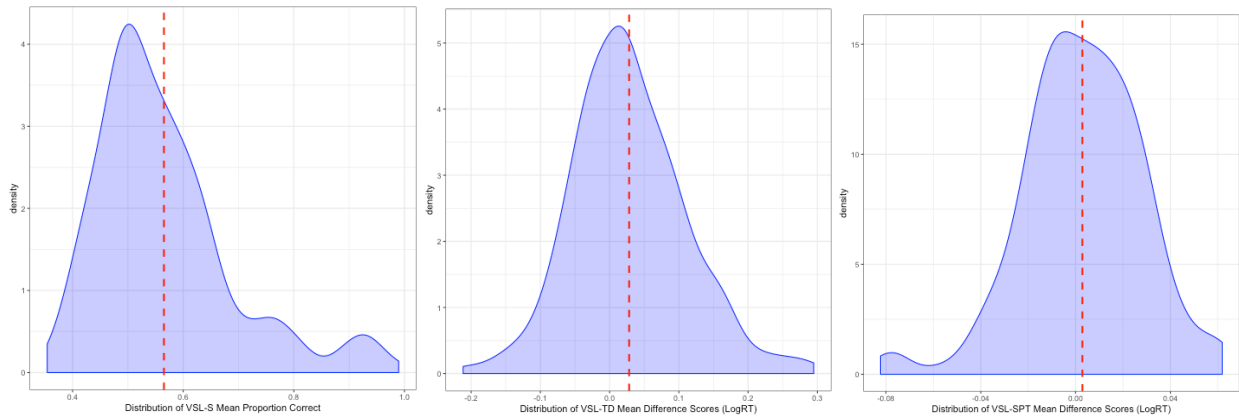
- Rose, M., Haider, H., Weiller, C., and Büchel, C. (2002). The Role of Medial Temporal Lobe Structures in Implicit Learning. *Neuron*, 36(6), 1221-1231. doi:10.1016/s0896-6273(02)01105-4
- Rueckl, J. (2010). Connectionism and the role of morphology in visual word recognition. *The mental lexicon*, 5(3), 371-400.
- Rueckl, J. G. (2016). Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading*, 20(1), 86-97. doi:10.1080/10888438.2015.1103741
- Rueckl, J. G., and Seidenberg, M. S. (2009). Computational modeling and the neural bases of reading and reading disorders. *How children learn to read: Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice*. (pp. 101-133) Psychology Press, New York, NY.
- Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W., Bick, A., Frost, S. J., Hancock, R., Wu, D. H., Mencl, W. E., Dunabeitia, J. A., Lee, J., Oliver, M., Zevin, J. D., Hoefft, F., Carreiras, M., Tzeng, O. J. L., Pugh, K. R., Frost, R. (2015). Universal brain signature of proficient reading: Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences*, 112(50), 15510-15515. doi:10.1073/pnas.1509321112
- Saffran, J. R. (2003). Statistical Language Learning. *Current Directions in Psychological Science*, 12(4), 110-114. doi:10.1111/1467-8721.01243
- Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical Learning by 8-Month-Old Infants. *Science*, 274(5294), 1926-1928. doi:10.1126/science.274.5294.1926
- Saffran, J. R., Newport, E. L., and Aslin, R. N. (1996). Word Segmentation: The Role of Distributional Cues. *Journal of Memory and Language*, 35(4), 606-621. doi:10.1006/jmla.1996.0032
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., and Barrueco, S. (1997). Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear. *Psychological Science*, 8(2), 101-105. doi:10.1111/j.1467-9280.1997.tb00690.x
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(3), 501-518. doi:10.1037//0278-7393.13.3.501
- Schapiro, A., and Turk-Browne, N. (2015). Statistical Learning. *Brain Mapping*, 501-506. doi:10.1016/b978-0-12-397025-1.00276-1
- Schendan, H. E., Searl, M. M., Melrose, R. J., and Stern, C. E. (2003). An fMRI Study of the Role of the Medial Temporal Lobe in Implicit and Explicit Sequence Learning. *Neuron*, 37(6), 1013-1025. doi:10.1016/s0896-6273(03)00123-5
- Schmalz, X., Altoè, G., and Mulatti, C. (2016). Statistical learning and dyslexia: a systematic review. *Annals of Dyslexia*, 67(2), 147-162. doi:10.1007/s11881-016-0136-0
- Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, B. Miller, J. R. Lee, and O. J. L. Tzeng (Eds.), *The extraordinary brain series. Dyslexia across languages: Orthography and the brain-gene-behavior link* (pp. 146-168). Baltimore, MD, US: Paul H Brookes Publishing.

- Seidenberg, M. S., and Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, 4(9), 353-361. doi:10.1016/s1364-6613(00)01515-1
- Seidenberg, M. S., and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523-568. doi:10.1037//0033-295x.96.4.523
- Shafro, C. L., Conway, C. M., Field, S. L., and Houston, D. M. (2012). Visual Sequence Learning in Infancy: Domain-General and Domain-Specific Associations With Language. *Infancy*, 17(3), 247-271. doi:10.1111/j.1532-7078.2011.00085.x
- Sherry, D. F., and Schacter, D. L. (1987). The evolution of multiple memory systems. *Psychological Review*, 94(4), 439-454. doi:10.1037//0033-295x.94.4.439
- Shohamy, D., and Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Sciences*, 14(10), 464-472. doi:10.1016/j.tics.2010.08.002
- Shohamy, D., and Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General*, 142(4), 1159-1170. doi:10.1037/a0034461
- Siegelman, N., and Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, 81, 105-120. doi:10.1016/j.jml.2015.02.001
- Siegelman, N., Bogaerts, L., and Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior research methods*, 49(2), 418-432.
- Siegelman, N., Bogaerts, L., Christiansen, M. H., and Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160059. doi:10.1098/rstb.2016.0059
- Simon, K. N., Werchan, D., Goldstein, M. R., Sweeney, L., Bootzin, R. R., Nadel, L., and Gómez, R. L. (2016). Sleep confers a benefit for retention of statistical language learning in 6.5 month old infants. *Brain and Language*, 167, 3-12. doi:10.1016/j.bandl.2016.05.002
- Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558-1568. doi:10.1016/j.cognition.2007.06.010
- Spencer, M., Kaschak, M. P., Jones, J. L., and Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and Writing*, 28(4), 467-490. doi:10.1007/s11145-014-9533-0
- Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of cognitive neuroscience*, 4(3), 232-243.
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171-177. doi:10.1016/j.nlm.2004.06.005
- Squire, L. R., and Knowlton, B. J. (1995). Memory, hippocampus, and brain systems. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 825-837). Cambridge, MA, US: The MIT Press.

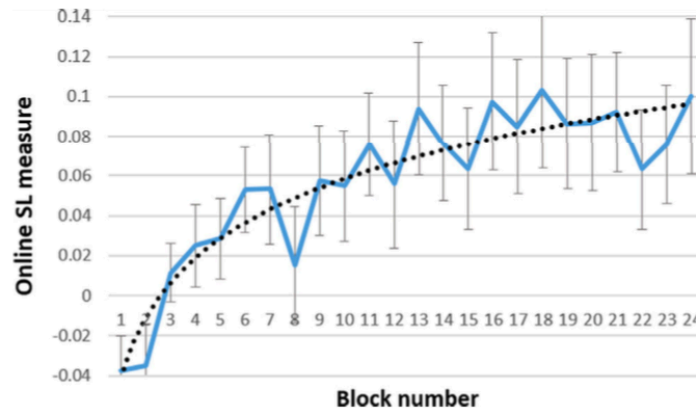
- Squire, L. R., and Knowlton, B. J. (2000). The medial temporal lobe, the hippocampus, and the memory systems of the brain. *The new cognitive neurosciences*, 2, 756-776.
- Stoodley, C. J., Harrison, E. P., and Stein, J. F. (2006). Implicit motor learning deficits in dyslexic adults. *Neuropsychologia*, 44(5), 795-798.
doi:10.1016/j.neuropsychologia.2005.07.009
- Stoodley, C. J., Ray, N. J., Jack, A., and Stein, J. F. (2008). Implicit Learning in Control, Dyslexic, and Garden-Variety Poor Readers. *Annals of the New York Academy of Sciences*, 1145(1), 173-183. doi:10.1196/annals.1416.003
- Strain, E., and Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 53(4), 347-359.
doi:10.1037/h0087322
- Suzuki, W. L., and Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: Cortical afferents. *The Journal of Comparative Neurology*, 350(4), 497-533. doi:10.1002/cne.903500402
- Taylor, J. S. H., Rastle, K., and Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin*, 139(4), 766-791.
doi:10.1037/a0030266
- Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160056.
doi:10.1098/rstb.2016.0056
- Thiessen, E. D., and Erickson, L. C. (2013). Beyond Word Segmentation. *Current Directions in Psychological Science*, 22(3), 239-243.
doi:10.1177/0963721413476035
- Thiessen, E. D., Kronstein, A. T., and Hufnagle, D. G. (2015). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin*, 139(4), 792-814. doi:10.1037/a0030801
- Treiman, R., Kessler, B., and Bick, S. (2003). Influence of consonantal context on the pronunciation of vowels: A comparison of human readers and computational models. *Cognition*, 88(1), 49-78. doi:10.1016/s0010-0277(03)00003-9
- Turk-Browne, N. B., Jungé, J. A., and Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology: General*, 134(4), 552-564.
doi:10.1037/0096-3445.134.4.552
- Turk-Browne, N. B., Scholl, B. J., Chun, M. M., and Johnson, M. K. (2009). Neural Evidence of Statistical Learning: Efficient Detection of Visual Regularities Without Awareness. *Journal of Cognitive Neuroscience*, 21(10), 1934-1945.
doi:10.1162/jocn.2009.21131
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92(1-2), 231-270.
doi:10.1016/j.cognition.2003.10.008
- Ullman, M. T., and Pierpont, E. I. (2005). Specific Language Impairment is not Specific to Language: the Procedural Deficit Hypothesis. *Cortex*, 41(3), 399-433.
doi:10.1016/s0010-9452(08)70276-4

- Vicari, S. (2005). Do children with developmental dyslexia have an implicit learning deficit? *Journal of Neurology, Neurosurgery and Psychiatry*, 76(10), 1392-1397. doi:10.1136/jnnp.2004.061093
- Vicari, S., Marotta, L., Menghini, D., Molinari, M., and Petrosini, L. (2003). Implicit learning deficit in children with developmental dyslexia. *Neuropsychologia*, 41(1), 108-114. doi:10.1016/s0028-3932(02)00082-9
- Voss, J. L., and Paller, K. A. (2008). Brain substrates of implicit and explicit memory: The importance of concurrently acquired neural signals of both memory types. *Neuropsychologia*, 46(13), 3021-3029. doi:10.1016/j.neuropsychologia.2008.07.010
- Wagner, A. D., Maril, A., and Schacter, D. L. (2000). Interactions between forms of memory: when priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, 12(Supplement 2), 52-60.
- Weinert, S. (2009). Implicit and explicit modes of learning: similarities and differences from a developmental perspective. *Linguistics*, 47(2). doi:10.1515/ling.2009.010
- Witt, A., Puspitawati, I., and Vinter, A. (2013). How Explicit and Implicit Test Instructions in an Implicit Learning Task Affect Performance. *PLoS ONE*, 8(1). doi:10.1371/journal.pone.0053296
- Woollams, A. M., Ralph, M. A., Madrid, G., and Patterson, K. E. (2016). Do You Read How I Read? Systematic Individual Differences in Semantic Reliance amongst Normal Readers. *Frontiers in Psychology*, 7. doi:10.3389/fpsyg.2016.01757
- Yang, J., and Li, P. (2012). Brain Networks of Explicit and Implicit Learning. *PLoS ONE*, 7(8). doi:10.1371/journal.pone.0042993
- Yang, Y., and Hong-Yan, B. (2011). Unilateral implicit motor learning deficit in developmental dyslexia. *International Journal of Psychology*, 46(1), 1-8. doi:10.1080/00207594.2010.509800
- Yi, H., Maddox, W. T., Mumford, J. A., and Chandrasekaran, B. (2014). The Role of Corticostriatal Systems in Speech Category Learning. *Cerebral Cortex*, 26(4), 1409-1420. doi:10.1093/cercor/bhu236
- Zhao, J., Wang, X., Frost, S. J., Sun, W., Fang, S., Mencl, W. E., Pugh, K. R., Shu, H., Rueckl, J. G. (2014). Neural division of labor in reading is constrained by culture: A training study of reading Chinese characters. *Cortex*, 53, 90-106. doi:10.1016/j.cortex.2014.01.003

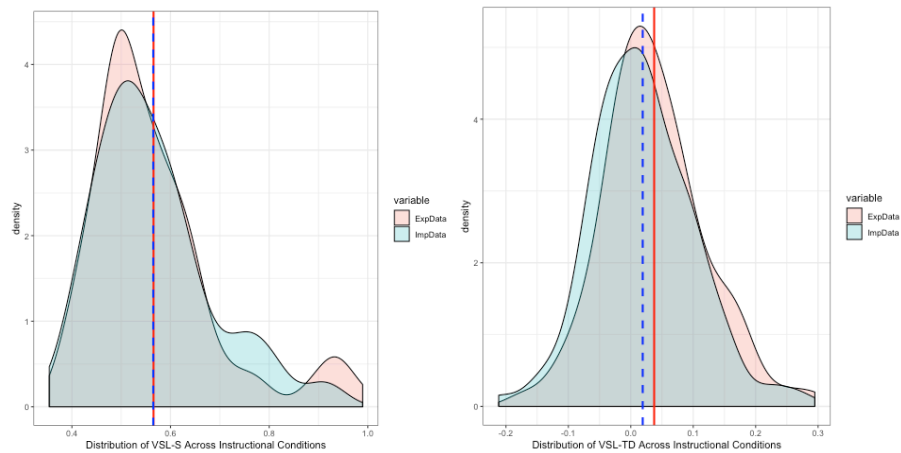
Appendix



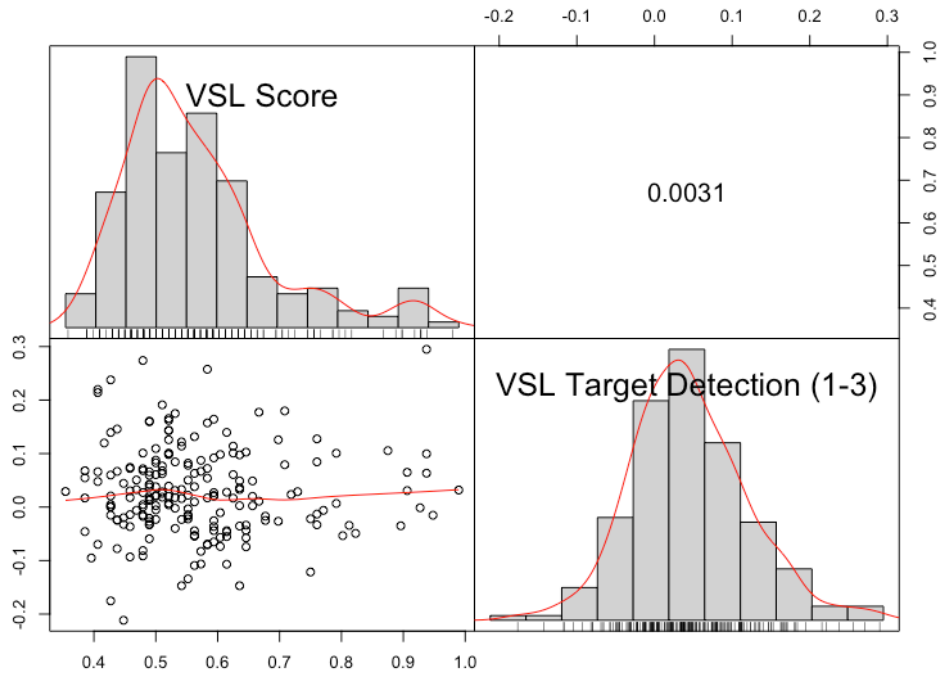
Density plots for VSL Score (VSL-Direct), VSL Target Detection (VSL-Indirect), and VSL Self Paced Task (VSL-SPT).



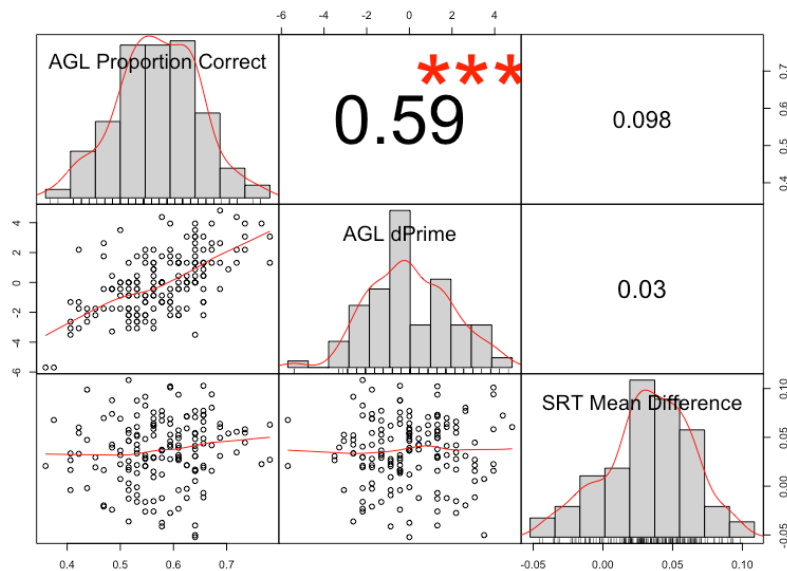
RT differences between predictable and unpredictable items (e.g., item 3 vs. item 1 in a triplet) in a similar self-paced VSL task from Siegelman et al. (2019). Unlike with the data presented in Figure 2, there is a clear increase in RT difference throughout the experiment. This suggests participants in their experiment the SPT was able to track the development of sensitivity to the statistical structure within the VSL.



Density plots of individual performance on VSL-Direct and VSL-Indirect split between instructional condition. Lines show respective mean values. VSL-Indirect differences in Log transformed units.



The distributions of VSL-Direct and VSL-Indirect are displayed. Scatterplots of these measures are also displayed with the Pearson's correlation. VSL-Direct is mean proportion correct. VSL-Indirect is a mean difference measure. RT is in Log.



Distributions, scatterplots, and correlations for Aim 2 IPM and EDM measures respectively. Scatterplots showing the relationship between items are displayed on the bottom left. Corresponding correlations and significance values displayed on the top right.

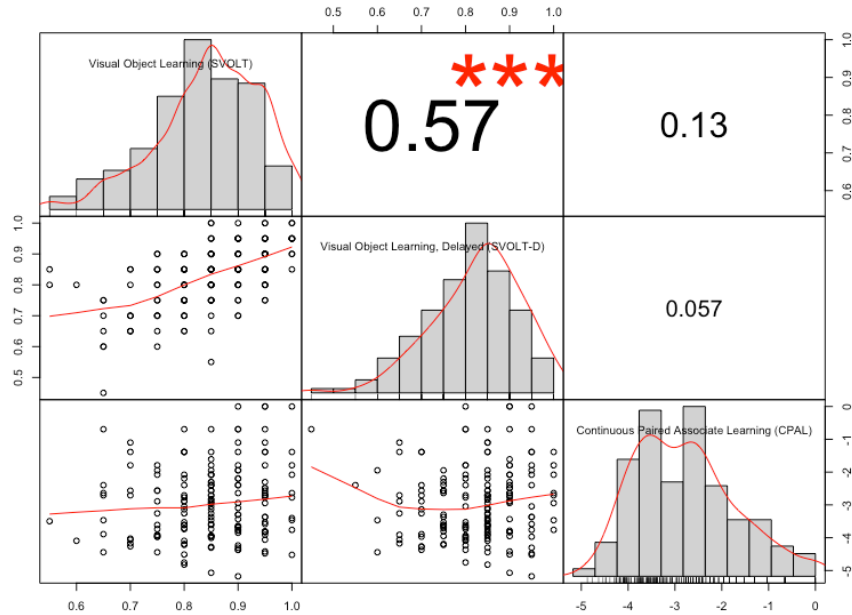


Figure 7. Distributions, scatterplots, and correlations for Aim 2 IPM and EDM measures respectively. Scatterplots showing the relationship between items are displayed on the bottom left. Corresponding correlations and significance values displayed on the top right.

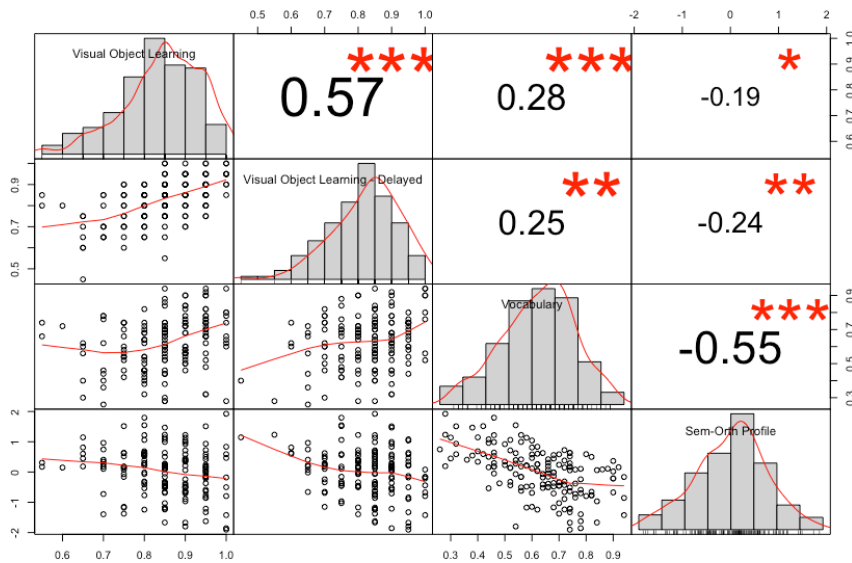


Figure 13. Distributions, scatterplots, and correlations for Aim 2 EDM and Language. Scatterplots showing the relationship between items are displayed on the bottom left. Corresponding correlations and significance values displayed on the top right.

Mean and Standard Deviation of VSL & IPM Measure Reliability (Pearson Correlation)	Model
--	-------

Construct	Measure	Subcomponent	Measurement/Analysis Type	Mean	SD	MEAN
VSL (Direct/Explicit), Statistical Control	VSL-Direct, All Data		Mean	0.73	0.02	Logit Transform of Mean Data
						glmer(ACC ~
			Logit	0.75	0.02	(1 target) + (1 participant), data=VSL.Split1, family=binomial(link=logit), glmerControl(optimizer = "bobyqa"))
			LME intercept	0.76	0.02	MEAN
	VSL-Direct, 2AFC Only		Mean	0.59	0.03	Logit Transform of Mean Data
						glmer(ACC ~ prevLocation +
			Logit	0.64	0.034	(1 target) + (1 participant), data=VSL.Split1, family=binomial(link=logit), glmerControl(optimizer = "bobyqa"))
						glm(ACC ~ prevLocation , data=temp2, family=binomial(link="logit"))
			LME Intercept	0.61	0.03	
			Ind Reg Intercept	0.25	0.1	MEAN
VSL (Indirect/Implicit), Statistical Control	VSL-Indirect					lmer(LogRT ~ Location + prevLocation + Trial.SD + LogPrevRT + Item + (1 SearchTriplet) + (1 + Item + presentation.order + trial participant), data=VSL.r.Split1)
			Difference	0.1	0.05	

		LME Coefficient		0.55	0.16	lm(LogRT ~ Location + prevLocation + Trial.SD + LogPrevRT + Item , data = temp2)
		Ind Reg Coefficient		0.09	0.06	MEAN
SRT, Blocks 3-8		Difference		0.22	0.06	MEAN
		Difference		0.26	0.06	lmer(LogRT ~ Prob3 + block + trial.SD + block*trial.SD + prevTrial + PrevRT + (1 stim) + (1 + Prob3 + trial.SD + prevTrial participant), data=SRT.Split1)
Implicit/Procedural (IPM), Statistical Control	SRT, ALL					
		LME Coefficient		0.52	0.09	lm(LogRT ~ Prob3 + block + trial.SD + block*trial.SD + PrevRT + , data = temp2)
		Ind Reg Coefficient		0.3	0.05	dPrime conversion
Grammaticality	dPrime			0.4	0.07	MEAN

							<pre> AGL.2.s1 = glmer(ACC ~ trial.SD + PrevGram2 + PrevClust2 + Clustering (1 participant), data=AGL.Split1, family=binomial(link=logit), glmerControl(optimizer = "bobyqa")) </pre>
			Mean		0.28	0.05	
Implicit/Procedural (IPM), Statistical Control	AGL, Established (Grammaticality Judegment)						<pre> glmer(Endorse ~ trial.SD + PrevGram2 + Grammaticality2 + Clustering + Grammaticality2 * Clustering + (1 + Grammaticality2 + Clustering participant), data=AGL.Split1, family=binomial(link=logit), glmerControl(optimizer = "bobyqa")) </pre>
			LME Intercept		0.28	0.05	
<hr/>							
							<pre> glmer(Endorse ~ trial.SD + PrevGram2 + PrevClust2 + Grammaticality2 + Clustering + Grammaticality2 * Clustering + (1 + Grammaticality2 + Clustering participant), data=AGL.Split1, family=binomial(link=logit), glmerControl(optimizer = "bobyqa")) </pre>
Implicit/Procedural (IPM), Additional Measures	AGL, Endorsement	Grammaticality	LME Coefficient		0.35	0.05	

					<code>glm(Endorse ~</code> <code>trial.SD +</code> <code>PrevGram2 +</code> <code>PrevClust2 +</code> <code>Grammaticality2</code> <code>Clustering</code> <code>,</code> <code>data=temp2,</code> <code>family=binomial(link="logit"))</code>
Ind Reg Coefficient	0.19	0.11	+		
					<code>glm(Endorse ~</code> <code>trial.SD +</code> <code>PrevGram2 +</code> <code>PrevClust2 +</code> <code>Grammaticality2</code> <code>Clustering</code> <code>,</code> <code>data=temp2,</code> <code>family=binomial(link="logit"))</code>
LME Coefficient	0.34	0.06	+		
Clustering					<code>glmer(</code> <code>Endorse ~</code> <code>trial.SD +</code> <code>PrevGram2 +</code> <code>Clustering +</code> <code>(1 + Clustering</code> <code> participant),</code> <code>data=AGL.Split1.un,</code> <code>family=binomial(link=logit),</code> <code>glmerControl(optimizer =</code> <code>"bobyqa")</code> <code>)</code>
Ind Reg Coefficient	0.17	0.1			

				glm(Endorse ~	
				trial.SD +	
				PrevGram2 +	
				Clustering	
				,	
				data=temp2,	
				family=binomial(link="logit"))	
AGL, Endorsement, Ungram	Clustering	LME Coefficient	0.28	0.07	
		Ind Reg Coefficient	0.19	0.09	
